EARLY UNIVERSE COSMOLOGY AND ITS OBSERVATIONAL EFFECTS ON THE COSMIC MICROWAVE BACKGROUND

TOM CHARNOCK

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

THESIS SUBMITTED TO THE UNIVERSITY OF NOTTINGHAM

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

APRIL 2017

SUPERVISORS	PROFESSOR EDMUND COPELAND
	DOCTOR ADAM MOSS
EXAMINERS	PROFESSOR MARTIN KUNZ
	DOCTOR PAUL SAFFIN
SUBMITTED	31ST MARCH 2017
EXAMINED	6TH APRIL 2017
FINAL VERSION	7TH APRIL 2017

Abstract

This Thesis is written in three parts. The first part describes the analytic calculation of the unequal-time correlator of cosmic strings and superstrings. The first efficient constraint analysis of all string and superstring network parameters is performed. By studying the effect of cosmic strings on the cosmic microwave background (CMB) radiation it is discovered that cosmic strings must make up a vanishingly small proportion of the energy density of the universe. The constraints on string network parameters are all skewed toward reducing the magnitude of energy density arising from strings. Also in this Part, a better comprehension of the unconnected segment model (USM) was gained. In particular, a greater understanding of the string scaling parameter $L_{\rm f}$ was garnered, as well as finding the reason why the USM tends to provide greater power than simulations of Nambu-Goto cosmic strings.

The second part contains a detailed description of statistical cosmology and how differences between parameter constraints from different data sets can lead to misleading quantification of discordance. The majority of this part describes different methods of quantifying differences between probability distributions and how these can be interpreted. In particular, using the most up-to-date data possible, differences between parameter constraints using the CMB and probes of large scale structure (LSS) in the universe can be measured. With current data the discordance can be interpreted as a low level of disagreement, but the application of prior ranges on well known parameters can force the tension to be greater. Using data from earlier work, this issue is considered in greater detail, with extensions to the accepted ΛCDM model added to test if the discordance can be alleviated. These extensions include the addition of active or sterile neutrinos and even ad-hoc changes to the primordial power spectrum. Although there are slight hints that these may help, when considering only the new data it might be unwise to believe that the discordance between parameter distributions from different data sets exists to a degree where the modifications are necessary.

Finally, application of deep learning to astrophysical observations is discussed. Using neural networks to learn about specific problems is *de rigueur* and their use in astronomy and cosmology is a promising field of study. In particular, applying raw data to neural networks can often outperform, or add enhanced features, to what is possible with current, non-empirical feature detection. The classification of supernovae from their light curves can be achieved using a specific machine learning architecture called a recurrent neural network (RNN).

Using the raw data from supernova light curves, the RNN is able to learn about features in sequences which can be used to classify types of supernova. Although a large training set is needed to perform as well as current techniques, one major advantage the RNN method has is the possibility of early detection. Rather than needing the entire light curve to perform statistical fits to categorise the supernova type, relatively little information from the early observation data is needed to classify using the RNN. Installing RNN on machinery for observation would save a vast amount of time by early classification since only supernovae of interest can be concentrated on.

Declaration

No part of this Thesis has previously been submitted for a degree or other qualification at this or any other university.

This Thesis is based on the following research:

Tom Charnock, Anastasios Avgoustidis, Edmund J. Copeland and Adam Moss. CMB constraints on cosmic strings and superstrings. Physical Review D, 93 (12) 123503, 2016.

Richard A. Battye, Tom Charnock, and Adam Moss. Tension between the power spectrum of density perturbations measured on large and small scales. Physical Review D, 91 (10) 103508, 2015.

Tom Charnock, Richard A. Battye and Adam Moss. Planck confronts large scale structure: methods to quantify discordance. Submitted to Physical Review D, arXiv:1703.05959, 2017.

Tom Charnock and Adam Moss. Deep recurrent neural networks for supernovae classification. The Astrophysical Journal Letters, 837 L28, 2017.

Contents

I Introduction **1**

1 Theory **6**
1.1 Manifolds, curvature and parallel transport 6

2 ΛCDM cosmology **10**
2.1 Friedmann equation and cosmological evolution 10
2.2 Cosmic microwave background 13
2.2.1 CMB temperature 13
2.2.2 Anisotropies 13
2.3 Power spectrum 15

3 Calculating observable effects **17**
3.1 Einstein-Boltzmann equations 17
3.2 Power spectrum shape 18
3.3 Cosmological parameters 20
3.4 Markov chain Monte Carlo 23

II Cosmic strings **27**

1 Cosmic strings **30**
1.1 Spontaneous symmetry breaking 30
1.2 Size and tension of a string 32
1.3 String evolution 33
1.4 Observational signatures from cosmic strings 35
1.5 Other topological defects 37

2 Cosmic string spectra **39**
2.1 Unequal-time correlator 39
2.1.1 String energy-momentum tensor 39
2.1.2 Velocity dependent one-scale model 41
2.1.3 Unconnected segment model 44
2.1.4 Analytic calculation of the unequal-time correlator 47
2.1.5 Eigenmode decomposition 50
2.1.6 Comparison of the string power spectrum 54
2.2 Cosmic superstrings 56
2.3 Cosmic string constraints 60

3 Discussion **68**

A Integral identities and analytic coefficients **71**

III Statistical cosmology **74**

1 Quantifying discordance **78**
1.1 Methods to quantify discordance 78

2 Cosmological parameter constraints **84**
2.1 Cosmic microwave background 84
2.2 Large scale structure 86
2.2.1 Weak gravitational lensing 87
2.2.2 Redshift-space distortions 91
2.2.3 Sunyaev–Zel'dovich galaxy cluster counts 94
2.3 Parameter constraints 97
2.4 ΛCDM extensions 106
2.4.1 Neutrinos 106
2.4.2 Alternative explanations 119

3 Discussion **128**

B Comparison of methods **131**

IV Deep learning **142**

1 Machine learning **144**
1.1 Perceptrons 144
1.2 Artificial neural networks 145
1.3 Back propagation 146
1.4 Activation functions 149
1.5 Recurrent neural networks 151
1.6 Data augmentation 154
1.7 Overfitting 155

2 Supernovae classification **157**
2.1 Example data 159
2.2 Network architecture 163
2.3 Results 164

3 Discussion **170**

List of figures

I Introduction **2**

I–3.1 *Planck*2015 temperature power spectrum 18

II Cosmic strings **28**

II–1.1 Manifold of allowed field values 31
II–1.2 Cosmic string formation 33
II–1.3 Observational signatures of cosmic strings 35

II–2.1 Cosmic string velocity and correlation length evolution . . . 43
II–2.2 String C_ℓ for scaling parameter $L_\mathrm{f} = 0.5,\ 1.$ 46
II–2.3 Analytically approximated UETC regions 49
II–2.4 Cosmic string UETC 51
II–2.5 UETC oscillations 52
II–2.6 C_ℓ error ratios 53
II–2.7 String C_ℓ comparisons 54
II–2.8 Cosmic superstring velocity and correlation length evolution 60
II–2.9 Cosmic string C_ℓ for various c_r and α 61
II–2.10 Cosmic superstring C_ℓ for various g_s, c_s and w 63
II–2.11 Cosmic string parameter constraints 65
II–2.12 Cosmic superstring parameter constraints 67

III Statistical cosmology **75**

III–2.1 Quadrupole component of the RSD power spectrum 94
III–2.2 $\Omega_\mathrm{m} - \sigma_8$ contours, LSS and CMB (2013) 96
III–2.3 $\Omega_\mathrm{m} - \sigma_8$ contours, LSS and CMB (2015) 97
III–2.4 $\Omega_\mathrm{m} - \sigma_8$ contours, *All LSS* and CMB (2015) 100
III–2.5 2D contours for all ΛCDM parameters (*All LSS (Weak)*) . . 105
III–2.6 2D contours for all ΛCDM parameters (*All LSS (Strong)*) . 106
III–2.7 $\Omega_\mathrm{m} - \sigma_8$ contours, *All LSS* and CMB in ΛCDM+$\sum m_\nu$ . . . 108
III–2.8 $\sum m_\nu$ values, CMB + range of LSS (2013) 109
III–2.9 $\sum m_\nu$ likelihoods, CMB + range of LSS (2013) 111
III–2.10 Bayes factor for likelihood compared to σ 112
III–2.11 Masses of the neutrino eigenstates 113
III–2.12 $\Omega_\mathrm{m} - \sigma_8$ contours, *All LSS* and CMB with $m^\mathrm{eff}_\mathrm{sterile}$+$\Delta N_\mathrm{eff}$ . . 117

III–2.13 $m_{\mathrm{eff}}^{\mathrm{sterile}} - \Delta N_{\mathrm{eff}}$ likelihoods, CMB + range of LSS (2013) . . 117
III–2.14 Matter power spectra for extended ΛCDM 121
III–2.15 CMB temperature power spectra for extended ΛCDM 122
III–2.16 τ_{R} likelihoods, CMB and CMB + LSS (2013) 125

III–B.1 Comparison of identical distributions 139
III–B.2 Comparison of distributions with different variances 139
III–B.3 Comparison of discordant probability distributions 140
III–B.4 Comparison of slightly shifted distributions 140
III–B.5 Comparison of unusually shaped distributions 141

IV Deep learning **143**

IV–1.1 Artificial neuron 144
IV–1.2 Fully connected artificial neural network 146
IV–1.3 Example activation functions 151
IV–1.4 LSTM cell 152

IV–2.1 Bidirectional recurrent neural network 158
IV–2.2 SN light curve and type-Ia probability 160
IV–2.3 Distribution of observation time for a given array length . . 160
IV–2.4 Training and test losses and accuracies 164

List of tables

I Introduction **2**

I–3.1 *Planck*2015+low-ℓ ΛCDM parameter constraints 22

II Cosmic strings **28**

II–A.1 Amplitude coefficients for the UETC 72
II–A.2 Approximation coefficients for the UETC and ETC 73

III Statistical cosmology **75**

III–2.1 Discordance, CMB and range of LSS (2015) 99
III–2.2 Discordance comparison, CMB and *All LSS (Weak)* 100
III–2.3 Discordance comparison, CMB and *All LSS (Strong)* 101
III–2.4 Parameter values, ΛCDM+$\sum m_\nu$ 110
III–2.5 Parameter values, ΛCDM+$m^{\rm eff}_{\rm sterile} + \Delta N_{\rm eff}$ 118
III–2.6 Parameter values, ΛCDM+$\sum m_\nu + A_{\rm L}$ 123
III–2.7 Parameter values, ΛCDM+$m^{\rm eff}_{\rm sterile} + \Delta N_{\rm eff} + A_{\rm L}$ 124
III–2.8 Parameter values, ΛCDM+$\tau_{\rm R}$ 125
III–2.9 Parameter values, ΛCDM+$\tau_{\rm R} + A_{\rm L}$ 126

III–B.1 1D probability distributions being compared 131
III–B.2 2D probability distributions being compared 132

IV Deep learning **143**

IV–2.1 Data augmentation 161
IV–2.2 Results for type-Ia vs. non-type-Ia classification 166
IV–2.3 Results for 123 and early epoch SN1 and 123 classification . 169

Conventions

A few points on convention in this Thesis should be mentioned. Throughout, a mostly positive metric $(-,+,+,+)$ will be used, with the speed of light $c = 1$ always. Greek subscripts label spacetime coordinates and the Einstein summation convention will be used as in Part I–1. Bold parameters indicate vectors over spacetime dimensions.

Unit	Name	Variant	Measure
$^\circ$	Degree		Angle
arcmin	Minute of arc	$(1/60)^\circ$	Angle
eV	ElectronVolt		Energy
GeV	GigaelectronVolt	$\times 10^9$eV	Energy
K	Kelvin		Temperature
μK	Microkelvin	$\times 10^{-6}$K	Temperature
m	Metre		Length
cm	Centimetre	$\times 10^{-2}$m	Length
km	Kilometre	$\times 10^3$m	Length
pc	Parsec		Length
Mpc	Megaparsec	$\times 10^6$pc	Length
s	Second		Time

Units used in this Thesis

Acronym	Name	Page	Parts
123	Supernova type-I, -II, -III classification	162	IV
ACT	Atacama Cosmology Telescope	84	III
ANN	Artificial neural network	145	IV
AUC	Area under the curve	163	IV
BAO	Baryon acoustic oscillations	19	I, III
BB	*B*-mode autocorrelation	62	II, III
BK*Planck*	*Planck*/BICEP2/Keck cross spectra	62	II
BOSS	Baryon Oscillation Spectroscopic Survey	86	III
CDM	Cold dark matter	2	I, III
CFHTLenS	Canada France Hawaii Lensing Survey	86	III
CMB	Cosmic microwave background	2	I – III
COBE	Cosmic Background Explorer	2	I
DES	Dark Energy Survey	90	III, IV
EE	*E*-mode autocorrelation	62	II, III

Acronym	Name	Page	Parts
FLRW	Friedmann-Lemaître-Robertson-Walker	10	I, II
GRU	Gated recurrent unit	152	IV
GUT	Grand unified theory	3	I, II
HFI	High Frequency Instrument (*Planck*)	3	I – III
highL	ACT+SPT high ℓ TT	84	III
ISW	Integrated Sachs-Wolfe effect	18	I, III
LFI	Low Frequency Instrument (*Planck*)	3	I – III
`lollipop`	Low ℓ EE polarisation likelihood (2016)	22	I, III
lowP	Low ℓ TE, EE and BB	62	II, III
LSS	Large scale structure	4	I, III
LSST	Large Synoptic Survey Telescope	36	II, IV
LSTM	Long short term memory	152	IV
MCMC	Markov chain Monte Carlo	23	I – III
NGC	North galactic cap	94	III
Pol	HFI EE and TE and LFI EE, TE and BB	85	III
RMS	Root-mean-square	41	I – IV
RNN	Recurrent neural network	146	IV
RSD	Redshift-space distortions	4	I, III
SDSS	Sloan Digital Sky Survey	85	III
SGC	South galactic cap	94	III
SN1a	Supernova type-Ia vs. non-Ia	162	IV
SPCC	Supernovae Photometric Classification Challenge	143	IV
SPT	South Pole Telescope	84	III
SZ	Sunyaev–Zel'dovich	3	I, III
TE	Temperature and E-mode cross correlation	62	II, III
TT	Temperature autocorrelation	62	II, III
UETC	Unequal-time correlator	29	II
USM	Unconnected segment model	29	II
VOS	Velocity dependent one-scale model	41	II
WMAP	Wilkinson Microwave Anisotropy Probe	2	I, III
WP	WMAP polarisation	85	III
WtG	Weighing the Giants	96	III
ΛCDM	Standard cosmological model	2	I – III

Acronyms used in this Thesis

Symbol	Description	Page	Parts
$a(\tau)$	Scale factor (also $a(t)$)	10	I – III
$a_{\ell m}$	Complex coefficients for spherical harmonics	14	I
$A(\theta)$	Integration bounds	80	III
$A_i(x_\pm, \varrho)$	Analytic string amplitudes (table II–A.1)	47	II
A_{L}	*Planck* lensing parameter	122	III
A_{s}	Amplitude of scalar perturbations	20, 21	I, III
$A_\mu(x)$	Vector gauge boson	30	II
$1-b$	Galaxy mass bias	95	III
b_j^l	Bias of node j	144	IV
$B(\alpha, x_i)$	Small $x_{1,2}$ string UETC amplitudes	48	II
$B(\theta)$	Bhattacharyya distance	78	III
$c_i(w, g_{\mathrm{s}})$	String self-interaction coefficient	58	II
c_{r}	String loop chopping efficiency	29	II
c_{s}	Superstring self-interaction coefficient	29	II
$C(\alpha, x_i)$	Cosmic string equal time correlator	48	II
$C(\theta)$	Difference vector coefficient	79	III
C_ℓ	Power spectrum amplitude	15	I – III
$\mathbf{C}_t$	Cell state at time t	153	IV
$d_{ij}^k(w, g_{\mathrm{s}})$	String cross-interaction coefficient	58	II
D_{A}	Angular diameter distance	85	III
D_{V}	Comoving volume distance	23, 85	I, III
$D_{(i)}$	Data label	23	I, III
$D(P_1 \| \| P_2)$	Relative entropy (KL-divergence)	81	III
$\langle D \rangle(\theta)$	Expected entropy	81	III
f	Growth of structure parameter	23, 92	I, III
$f(\xi, \tau)$	String network scaling function	45	II
F_1	SPCC figure of merit	163	IV
F_{AP}	Alcock-Paczynski factor	23, 92	I, III
g_{s}	String coupling constant	29	II
$g_{\mu\nu}(x)$	Components of the metric tensor g	7	I
G	Newton's constant	10	I, II
$(G)\mu$	(Dimensionless) cosmic string tension	28	II
$(G)\mu_{\mathrm{F}}$	(Dimensionless) cosmic superstring tension	29	II
$\mathcal{G}_{\mu\nu}(x)$	Components of the Einstein tensor	9	I
h	Dimensionless Hubble parameter	20	I, III
H_0	Value of the Hubble function today	20	I – III

Symbol	Description	Page	Parts
$H(t)$	Hubble function	85	III
$\mathcal{H}(\tau)$	Hubble function in conformal units	12	I, II
$[h_i]$	Dense connections in hidden layers	163	IV
$\mathbf{h}_t$	Hidden state vector at time t	152	IV
$I_i(x_\pm, \varrho)$	Analytic string integral identities (II–A)	47	II
$I_i(\theta)$	Integral between interval integration value	80	III
k	Magnitude of wavevector $\mathbf{k}$	15	I–III
$\tilde{k}(v)$	Description of small-scale structure on string	41	II
ℓ	Multipole describing functions on a sphere	14	I – III
$L(\tau)$	Length of string segment in the network	34	II
$L(w^l_{ji}, b^l_j)$	Loss function of neural network	147	IV
L_f	String scaling parameter	44	II
m_ν	Neutrino mass eigenstate	113	III
$\mathcal{M}$	Example model	23	I, III
$\hat{n}$	Unit vector describing direction on the sky	14	I – III
n_s	Spectral tilt of power spectrum	20, 21	I, III
$\mathcal{N}(\mu, \sigma)$	Gaussian with mean μ and variance σ	101	III
$O(\theta)$	Overlap coefficient	79	III
$p(\theta)$	Prior on parameters θ	23	I, III
$p(D)$	Evidence of data D	23	I, III
$p(D_1, D_2)$	Joint evidence of data D_1 and D_2	82	III
$P(\theta\|D, \mathcal{M})$	Posterior distribution of parameters θ	23	I, III
$P_{(0\|i)}(k)$	Primordial power spectrum	16, 21	I, III
$P_\ell(\hat{n} \cdot \hat{n}')$	Legendre polynomial for multipole ℓ	15	I, III
$P_\mathrm{g}(k)$	Galaxy power spectrum	91	III
$r_{\mathrm{HD_A}}$	Correlation radius between H and D_A	86	III
$r_\mathrm{s}(z_\mathrm{d})$	Distance to sound horizon at the drag epoch	23, 85	I, III
$R(\theta)$	Bayesian evidence ratio	81	III
$S(\theta)$	Surprise ($D(P_1\|\|P_2) - \langle D \rangle$)	81	III
$T(\theta)$	Shifted probability distribution evidence ratio	82	III
$T_{\mu\nu}(x)$	Components of the energy-momentum tensor	11	I, III
T_C	Symmetry breaking critical temperature	30	II
u_i	Eigenvector of diagonalised UETC	50	II
$v_{(i)}$	String velocity (of string i)	40	II
v^l_j	Weighted, biased neuron input	144	IV
$V(\|\varphi\|)$	Abelian-Higgs potential	30	II

Symbol	Description	Page	Parts
w	Equation of state parameter	11	I
w	Effective volume of extra compact dimensions	29	II
w_{ji}^{l}	Network weight between node i and node j	144	IV
$x_i(k, \xi_i, \tau_i)$	String $k\xi_i\tau_i$	47	II
x_i	Network input vector	144	IV
$x_{\pm}(x_i)$	Average and difference x_i $((x_1 \pm x_2)/2)$	47	II
Y_{ℓ}^{m}	Spherical harmonics	14	I, III
z_{eff}	Effective redshift (eff = CMASS, LOWZ, etc.)	85	III
α	String wiggliness parameter	29	II
$\alpha_{\parallel}$	Scaling factor along the line-of-sight	85	III
$\alpha_{\perp}$	Scaling factor perpendicular to line-of-sight	85	III
$\delta_{\alpha\beta}$	Krönecker δ	9	I, III
$\delta^{(d)}$	d dimensional Dirac δ	16	I, II
Δ	String deficit angle $(8\pi G\mu)$	35	II
ΔN_{eff}	Additional effective degrees of freedom	22, 115	I, III
γ	Cosmological shear	87	III
η	Symmetry breaking scale	30	II
η	Machine learning rate	148	IV
κ	Spatial curvature parameter	10	I
κ	Superstring cross-interaction coefficient	58	II
κ	Weak lensing convergence	87	III
λ	Abelian quadratic coupling constant	30	II
λ_i	Eigenvalues of diagonalised UETC	50	II
Λ	Cosmological constant	10	I
μ	Mean value of distribution	79	III
ξ	String correlation length	32	II
$\varrho(k, v_i, \tau_i)$	String $k\lvert v(\tau_1)\tau_1 - v(\tau_2)\tau_2\rvert$	47	II
$\phi(v_j^l)$	Activation function	144	IV
$\varphi(x)$	Complex scalar field	30	II
ψ	Model extension parameters	26	I, III
$\psi(x)$	Gravitation lensing potential	87	III
$\varrho_i(\tau)$	Energy-density of a perfect fluid	11	I
σ_8	RMS perturbation in $8h^{-1}$Mpc spheres	23	I, III
$\sigma_{\text{i}}(x_i)$	Sigmoid function	153	IV
τ	Conformal (or comoving) time	10	I, II
τ_{R}	Optical depth to reionisation	20, 21	I, III

Symbol	Description	Page	Parts
θ	Parameter vector	23	I, III
$\vartheta(x)$	Phase of complex scalar field	31	II
$\Theta(\hat{n})$	Temperature anisotropy	14	I – III
$\Theta^{(SVT)}_{(\mu\nu)}(x)$	Energy-momentum tensor of strings	39	II
Θ_{MC}	Size of the sound horizon at last scattering	20	I, III
Θ_{B}	Bayes factor	26	I, III
Ω_i	Dimensionless density parameter ($\varrho_i/\varrho_{\mathrm{c}}$)	12	I, III
$\Omega_{\mathrm{b}}h^2$	Physical density of baryons	20	I–III
$\Omega_{\mathrm{c}}h^2$	Physical density of cold dark matter	20	I–III
Ω_{m}	Fraction of the matter energy-density	12	I, III
$\sum m_\nu$	Sum of the neutrino mass eigenstates	22, 107	I, III
∂_μ	Partial derivative with respect to x^μ ($\partial/\partial x^\mu$)	8	I – IV
$'$	Derivative with respect to τ ($\partial/\partial\tau$)	12	I, II
∇_μ	Covariant derivative along basis vector $\mathbf{e}_\mu$	8	I – III
$\langle\ \rangle$	Ensemble average over infinite realisations	14	I – III
$\odot$	Hadamard product	147	IV

Notation used in this Thesis

Part I

Introduction

Never before have such accurate measurements of the evolution of the universe been available. Using the incredible amounts of accumulated data on cosmic history it is now possible to interpret how the cosmos must have matured into the domain that currently exists. The field of modern cosmology began almost exactly a century ago with the advent of Einstein's general relativity [124] followed by Lemaître's and Friedmann's prediction of the expansion of the universe [135; 231], confirmed by Hubble in 1929 [188]. This was furthered by the conjecture of the cosmic microwave background (CMB) radiation (described in Part I–2.2) by Gamow, Alpher and Hermann [26; 136] and its vindication in 1965 by Penzias and Wilson [284]. Precision cosmology, in truth, began with the Cosmic Background Explorer (COBE) satellite and its measurements of the thermal spectrum of the CMB from 1989-1993 [254]. From this time onwards observations of the cosmos have become progressively detailed, pinning cosmological features such as the amount of matter and energy, and the expansion history. From this, it can be inferred that the universe probably began with a period of inflation, settling into expansion driven by, firstly radiation pressure, followed by a matter era and finally entering a stage of dark energy domination [122; 152]. Using the precision measurements available, the total energy density in the universe coming from the everyday matter that we see around us must account for ~5%. Around 25% of the energy density comes from a form of weakly interacting "dark matter" which must be non-relativistic, or "cold", to fit current observations. This is known as cold dark matter (CDM). The remaining 70% of the energy budget exists as "dark energy", a mysterious force - still not well understood - which causes the expansion of the universe to accelerate [343]. More about this standard model of cosmology, referred to as ΛCDM, will be presented in Part I–2.

CMB observations reached their pinnacle with the *Planck* experiment [354]. This is the third generation CMB mission following the aforementioned COBE satellite and the Wilkinson Microwave Anisotropy Probe (WMAP) [53]. COBE's Far-InfraRed Absolute Spectrophotometer (FIRAS) first measured the blackbody spectrum of the CMB at $T_{\mathrm{CMB}} = 2.73K$ [254]. It also showed that anisotropies existed only on a level of one part in one-hundred-thousand, i.e. the anisotropies had an amplitude of $\Delta T \lesssim 10\mu$K [253]. COBE's successor, WMAP managed to measure these anisotropies to such an extent that it could constrain the age of the universe to 13.8 billion years within a precision of 1% [53]. WMAP was also instrumental as placing ΛCDM as the standard model of cosmology, measuring the distribution of energy density between radiation, baryonic matter, CDM and dark energy [53]. WMAP's finding suggested that the geometry of the universe was consistent with being flat with an initial period of inflation [53]. As well as

measuring the temperature anisotropies of the CMB, WMAP detected the E-mode polarisation - a curl-free signature of photons scattering from electrons [53]. The results from the polarisation measurements suggested that primordial density fluctuations were adiabatic, i.e. an underdensity in one species, say the photon, would be accompanied by an underdensity in other species [53]. The *Planck* satellite advanced WMAP's achievements, measuring the CMB anisotropies at three times higher resolution, making the measurements limited only by foreground subtraction and not by the detector performance [354]. The satellite contained two instruments, a Low Frequency Instrument (LFI) and a High Frequency Instrument (HFI), both capable of measuring the temperature and polarisation of photons. The detailed measurements from *Planck* secured ΛCDM's position and suggested that extensions, such as massive neutrinos or modified gravity were not essential or even likely [14]. Not only was *Planck* capable of observing the CMB anisotropies and its polarisation but it also had a wide variety of other scientific objectives. Two worth mentioning here are the surveying of weak gravitational lensing of the CMB - the perturbation of CMB photon trajectories by the gravitational potential around dense objects such as galaxy clusters [15], and the counting of the number of galaxy clusters using the Sunyaev–Zel'dovich (SZ) effect - the heating of photon temperature by inverse Compton scattering with high energy electrons in the intracluster plasma [17].

Since measurements of the universe have become so specific, theoretical predictions about what may exist beyond ΛCDM can be tested. Although observations from *Planck* suggest there is no distinguishable deviation from ΛCDM there is still some scope for constraints to be placed on extensions of the standard cosmological model. Features in the CMB give an indication of physics which occurred earlier in the history of the universe than the surface where photons last scattered with ionised protons and electrons, a time known as recombination [281; 386]. Before this era, the universe was at an extremely high temperature - high enough that we know at least the electromagnetic and weak forces were unified [145; 315; 374; 375] and that the strong force could be unified under a larger grand unified theory (GUT) [140; 278]. Signatures of a GUT could be left behind during symmetry-breaking phase transitions [103]. The possibility of these signatures from topological defects, in particular cosmic strings, being detected is discussed in Part II. Here, the two-point correlation function of cosmic strings is analytically calculated, allowing constraints on the string network parameters to be obtained using observations of the CMB.

Although the CMB as measured by *Planck* is in complete agreement with the

theoretical predictions of a ΛCDM cosmology, this does not take into account what other data may indicate about the state of the universe. It has been noticed that there are a variety of probes of large scale structure (LSS) with parameter constraints that do not agree with the constraints obtained using *Planck* in a ΛCDM model [5; 12; 45; 57; 112; 138; 156; 183; 222; 245; 300; 310; 382]. This discordance suggests that there is a possibility of an extension to the standard cosmological model to bridge the differing results obtained on the largest scales of the CMB to the smaller scales of LSS. To calculate whether the discordance exists, a deep understanding of the differences between high dimensional cosmological probability distributions is necessary. Part III–1 introduces several measures which quantify the similarity or difference between posterior distributions with the same parameters in the same model, but constrained by different data sets. Once a measure is established, the constraints from cosmological data can be considered. Loosely, measurements of LSS prefer less small scale structure than CMB observations. This is noticed for a number of different observational probes including: redshift-space distortions (RSD) [58] - the squashing of galaxies in real-space due to their observation in redshift-space; weak gravitational lensing [10; 169] - the warping of a photon's path by the variation in gravitational potential it passes through; and SZ galaxy cluster counts [12] - the number of galaxy clusters determined by the heating of photons due to inverse Compton scattering in the intracluster plasma. Details of these probes can be found in Part III–2.2. Since each of these probes are independent, non-correlated observations, they can be combined to form a much tighter set of parameter constraints than each probe independently. Nominally, it is seen that there is not a very statistically significant difference between the posterior distributions from the combined LSS data sets and *Planck*. The story changes when applying well known cosmological priors which shift the LSS posterior along correlated parameter directions into regions which are in greater tension with the *Planck* constraints. If these priors are to be believed then extensions to ΛCDM can be considered to alleviate the discordance. Some well motivated extensions (discussed in Part III–2.4) include the addition of active or sterile neutrinos and modifications to the primordial power spectrum.

The analysis of cosmological data requires ever more sophisticated computational techniques and greater understanding of assumed astrophysics. Future large-scale surveys will produce such large quantities of data, current processing techniques will not be powerful enough [107]. One of the, potentially promising, methods to improve data processing is the application of deep learning [39]. Deep learning is a way of generating a specific algorithm from a generic set of

expressions contained within “neurons”. Networks of these neurons are trained to take an input and interact such that the output is the required answer, and so provide a black box-like algorithm. In Part IV there is a description of machine learning and how neural networks can be trained. It is often necessary to pull the key components from data for analysis to be able to deal with large quantities of information [269]. In Part IV–2 the raw data from simulated supernova light curves are fed directly into a neural network. By learning directly from the data, without needing any fitting, early classification of supernovae can be achieved - well before the end of the whole observation. If techniques such as this were applied to current technology, such as telescopes, then huge amounts of observation time could be saved by focussing only on the desired supernovae type.

The remainder of this Part introduces a detailed description of ΛCDM and how it can be constrained via imprints in the CMB.

Chapter 1

Theory

To truly understand some aspects of this Thesis it will be useful to know some differential geometry terms such as manifold, covariant derivative, scalar, vector and tensor. To make it clear what these refer to, this chapter will contain a short description of some of the major points of general relativity which can be used as a guide throughout.

1.1 Manifolds, curvature and parallel transport

A manifold is a formally defined topological space, i.e. a set of points with neighbours which follow certain axioms [30; 264, Chapter 9]. More than just a set of points, a manifold is locally a Euclidean space - the n-dimensional space of real numbers $\mathbb{R}^n$ [30; 264, Chapter 9]. Maps into the n-dimensional space of complex numbers $\mathbb{C}^n$ can be conceived but are not considered here. Coordinates can now be defined since each point on the manifold has a one-to-one mapping into the space of real numbers [30; 264, Chapter 9]. Any point which is unique in $\mathbb{R}^n$ must therefore be unique on the manifold, although the manifold need not contain a map to all of $\mathbb{R}^n$ [30; 264, Chapter 9]. A subset of the points on the manifold can be mapped to a patch of coordinates in $\mathbb{R}^n$, which can be transformed to a different set by mapping back to the manifold and then forward to $\mathbb{R}^n$ using a different map [30; 264, Chapter 9].

Scalar, vector and tensor fields can be defined on this manifold. Firstly, a (real) scalar field is simply a function which assigns a point in the manifold to a number (or coordinate) in $\mathbb{R}^n$ [30; 264, Chapter 9]. At each point in the manifold there is a tangent vector space, a set of functions (called vectors) which act on the points on the manifold describing how a scalar function changes [30; 264, Chapter 9]. An example of this is the changing of a function in a timelike direction, i.e. a velocity. A vector field then is a selection of vectors from the vector space at

each point on the manifold showing how a monotonically (single-valued) increasing interval mapped into points on the manifold changes at each point [30; 264, Chapter 9]. This can then be interpreted as the direction of the flow of a function at each coordinate in the space of real numbers. One-forms are cotangent to vectors, meaning that a one-form acting on a vector (or a vector acting on a one-form) is a real number, i.e. a scalar field [30; 264, Chapter 9]. Finally, this can be generalised to a tensor field with rank (p, q) which describes a linear functional of p one-form fields and q vector fields [30; 264, Chapter 9].

Within this Thesis, particularly in this Part, the components of the metric tensor will be used to define the geometry of space. The metric tensor g, is a symmetric $(0, 2)$-type tensor field which acts on two vector fields to give a number in $\mathbb{R}^n$. This can be interpreted as the measure of the square of the length on the manifold, ds^2 [161, Chapter 7; 264, Chapter 11]. Coordinates in $\mathbb{R}^n$ are denoted x^μ, where greek indices label of the dimension n, of the manifold, i.e. $\mu = 0 \to n-1$. The distance relation between points on the manifold separated by infinitesimal coordinate intervals dx^μ is

$$ds^2 \quad = \quad g_{\mu\nu}(x) dx^\mu dx^\nu \ , \tag{I–1.1}$$

where $g_{\mu\nu}(x)$ are the components of the metric tensor in a given coordinate basis [109, Chapter 2; 161, Chapter 7; 264, Chapter 14]. Here similar indices are summed over, $\varsigma_\mu \varsigma^\mu \equiv \sum_{\mu=0}^{n-1} \varsigma_\mu \varsigma^\mu$. This notation will be used throughout the Thesis.

Manifolds need not be flat, and in fact, direction is conceptual [30; 264, Chapter 11]. Since vectors at different points are in different vector tangent spaces they cannot be compared, there is no idea of direction between tangent spaces. One way to introduce direction is to consider parallel transport via a covariant derivative [264, Chapter 14]. If there is a curve through two points on the manifold then the value of a vector at the second point from the first point's vector space will not necessarily be equal to the vector from the vector space at the second point. The covariant derivative is defined by comparing the vector from the first vector space, which is parallel to the curve at the second point, to the vector from the second vector space given that the points are infinitesimally separated [30; 264, Chapter 11]. Since this difference is equivalent to the vector at the first point being transported in a predefined way to the second point, it describes another vector. This is one of the useful properties of the covariant derivative, a derivative on a rank (p, q)-type tensor still has rank (p, q) [30; 264, Chapter 11]. Parallel transport requires vectors transform covariantly, the com-

ponents of a vector must change in a way such that the vector is invariant under change of coordinate basis. The corrections to the components of the vector are known as Christoffel symbols $\Gamma^{\alpha}{}_{\mu\nu}$, and ensure a change of basis vector, say from $\mathbf{e}_\mu$ to $\mathbf{e}_\nu$, define the same vector via

$$\nabla_\mu \mathbf{e}_\nu \quad = \quad \Gamma^{\alpha}{}_{\nu\mu}\mathbf{e}_\alpha \ , \tag{I–1.2}$$

where $\nabla_\mu \equiv \nabla_{\mathbf{e}_\mu}$ is the covariant derivative along the flow lines of the $\mathbf{e}_\mu$ basis vector [109, Chapter 2; 161, Chapter 22; 264, Chapter 11]. If a vector is parallel transported such that its basis vector doesn't change, then its covariant derivative vanishes. This allows the Christoffel symbols to be calculated from the components of the metric tensor. The metric tensor g, acts on vectors to preserve distances on the manifold at any point. Since the metric tensor acting on the two vectors is a real number, it's covariant derivative vanishes. This means the covariant derivative of the metric itself must vanish at every point. Using the definition of the covariant derivative acting on a rank $(0,2)$-type symmetric tensor, such as the metric, and knowing this vanishes allows the calculation of the Christoffel symbols

$$\Gamma^{\alpha}{}_{\mu\nu} \quad = \quad \frac{1}{2}g^{\alpha\beta}(\partial_\mu g_{\beta\nu} + \partial_\nu g_{\mu\beta} - \partial_\beta g_{\mu\nu}) \ , \tag{I–1.3}$$

where $g^{\alpha\beta}$ are the inverse components of the metric tensor $(g_{\mu\nu}(x))^{-1}$ with $\partial_\mu \varsigma \equiv \partial\varsigma/\partial x^\mu$ as differentiation with respect to the coordinate x^μ [109, Chapter 2; 161, Chapter 7; 264, Chapter 11]. The idea of curvature in the manifold can be constructed from the Christoffel symbols. Any non-zero displacement between a vector at a point on the loop and the vector parallel transported around the loop indicates curvature of the manifold [264, Chapter 11]. Requiring the vector and the transported vector be equal reveals a rank $(1,3)$-type tensor known as the Riemann curvature tensor. This parallel transporting of vectors about a curved manifold introduces the commutator of covariant derivatives along the direction of the initial vectors as well as the covariant derivative of the commutator of those vectors. The components of this tensor which preserve a change of coordinate basis (when the coordinate bases commute) are

$$R^{\alpha}{}_{\beta\mu\nu} \quad = \quad \partial_\mu \Gamma^{\alpha}{}_{\nu\beta} - \partial_\nu \Gamma^{\alpha}{}_{\mu\beta} + \Gamma^{\alpha}{}_{\mu\varrho}\Gamma^{\varrho}{}_{\nu\beta} - \Gamma^{\alpha}{}_{\nu\varrho}\Gamma^{\varrho}{}_{\mu\beta} \ . \tag{I–1.4}$$

These components anticommute under interchange of either the first two indices or the last two and commute under interchange of the first two with the last two [109, Chapter 2; 264, Chapter 11]. From the Riemann curvature tensor, two other important quantities for this Thesis can be calculated. These are the Ricci tensor and the Ricci scalar, both obtained via contraction of indices [109, Chapter 2; 264, Chapter 11]. Firstly the components of the Ricci tensor are

found by contracting the first and third indices $R_{\mu\nu} \equiv R^{\alpha}{}_{\mu\alpha\nu}$ and so can be calculated from

$$R_{\mu\nu} = \partial_\alpha \Gamma^{\alpha}{}_{\nu\mu} - \partial_\nu \Gamma^{\alpha}{}_{\alpha\mu} + \Gamma^{\alpha}{}_{\alpha\beta}\Gamma^{\beta}{}_{\nu\mu} - \Gamma^{\alpha}{}_{\nu\beta}\Gamma^{\beta}{}_{\alpha\mu} \,. \tag{I–1.5}$$

The Ricci scalar is finally found by contracting the inverse components of the metric tensor with the components of the Ricci tensor $R \equiv g^{\mu\nu}R_{\mu\nu}$ [109, Chapter 2; 264, Chapter 11]. The components of the metric tensor and its inverse can be used to raise and lower indices on vectors, one-forms and tensors such that $\varsigma^{\alpha}{}_{\nu} \equiv g^{\alpha\mu}\varsigma_{\mu\nu}$ or $g^{\mu\alpha}g_{\mu\beta} = \delta^{\alpha}_{\beta}$ where δ^{α}_{β} is the Krönecker δ, where $\delta^{\alpha}_{\beta} = 1$ when $\alpha = \beta$ but vanishes otherwise.

One final tensor should be described here, the components of the Einstein tensor $\mathcal{G}_{\mu\nu}$ [109, Chapter 2; 161, Chapter 7; 264, Chapter 14]. This contains the transverse components of the Ricci tensor and in four spacetime dimensions is

$$\mathcal{G}_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \,, \tag{I–1.6}$$

where, as a consequence of Bianchi identities, $\nabla^{\mu}\mathcal{G}_{\mu\nu} = 0$. It will be convention throughout to refer to the components of a tensor by the tensor's name, i.e. the components of the Einstein tensor is named the *Einstein tensor* and the components of the metric tensor called the *metric*.

Chapter 2

ΛCDM cosmology

2.1 Friedmann equation and cosmological evolution

The standard model of cosmology, named ΛCDM, depends on just a few key assumptions. It is presumed that the universe is both homogeneous and isotropic - the universe appears the same from any position with no preferred direction [135; 161, Chapter 18; 231]. One exact solution to Einstein's field equations, built on these assumptions, is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric [135; 231]. This describes a universe which can either be expanding or contracting, whose spatial expansion $a(t)$ can be time-dependent and is written

$$ds^2 = -dt^2 + a(t)^2\left[\frac{dr^2}{1-\kappa r^2} + r^2 d\mathbf{\Omega}^2\right], \tag{I–2.1}$$

where κ describes the curvature of space, dt and dr are the timelike and radial coordinates and $d\mathbf{\Omega}^2 = d\theta^2 + \sin^2\theta d\phi^2$ are spherical polar coordinates [161, Chapter 18]. κ can be normalised such that $\kappa = 1,\ 0,\ -1$ describes spherical, Euclidean and hyperbolic spaces respectively [161, Chapter 18]. κ will not be normalised when mentioned later so that more general statements about parameters can be made. It will be useful, throughout this Thesis, to consider the comoving - or conformal - time τ, which is related to physical time t via ([109, Chapter 2])

$$\tau \equiv \int \frac{dt}{a(t)}\,. \tag{I–2.2}$$

Einstein's field equations describe how energy affects the geometry of spacetime through ([109, Chapter 2; 161, Chapter 22])

$$\mathcal{G}_{\mu\nu} = 8\pi G T_{\mu\nu} - \Lambda g_{\mu\nu}\,, \tag{I–2.3}$$

where $\mathcal{G}_{\mu\nu}$ is the Einstein tensor from equation (I–1.6) and $g_{\mu\nu}$ is the metric. G is Newton's constant, Λ is the cosmological constant (further discussed below)

and $T_{\mu\nu}$ is the energy-momentum tensor which contains the information about the elements of matter in the universe. For perfect fluids, which will be solely discussed in this Thesis,

$$T_{\mu\nu} = (\varrho + p)u_\mu u_\nu + p g_{\mu\nu} \tag{I–2.4}$$

where $\varrho(t)$ is the energy density of each constituent form of matter and $p(t)$ is the corresponding pressure [161, Chapter 7]. $u_\mu \equiv g_{\mu\nu}u^\nu$ where u^ν is the four-velocity of the fluid where $g_{\mu\nu}u^\mu u^\nu = -1$.

There are three sources of energy to be concerned with in the universe: radiation; matter; and dark energy [109, Chapter 1]. Radiation describes relativistic species such as photons or massless (or even relativistic massive) neutrinos, whilst matter relates to the non-relativistic species such as baryons or CDM. Massive particles are relativistic when their kinetic energy is comparable to or greater than their rest energy [259, Chapter 3]. CDM is an essential ingredient to ΛCDM and is necessary to provide the correct structure growth in the universe (discussed in Part I–3.3). Dark energy is a force which causes the expansion of the universe to accelerate, the current cause of which is unknown [109, Chapter 1]. In ΛCDM cosmology the cosmological constant is responsible for dark energy. It provides an energy density which is constant through time [161, Chapter 18]. To see how each of the constituents affects cosmological expansion Einstein's field equations (given in equation (I–2.3)) are used along with an equation of state. The equation of state relates the pressure of one the constituents to its energy density, $p = w\varrho$ [161, Chapter 18; 376, Chapter 1]. For radiation and matter w can be found by comparing the energy density of particles in a given volume to the pressure within the volume. For radiation $w = 1/3$ and for matter $w = 0$ [161, Chapter 18; 376, Chapter 1]. The equation of state parameter for a cosmological constant can be found by absorbing the last term in equation (I–2.3) to the $T_{\mu\nu}$ term by transforming

$$\begin{aligned} \varrho &\rightarrow \varrho + \frac{\Lambda}{8\pi G}\,, \\ p &\rightarrow p - \frac{\Lambda}{8\pi G}\,. \end{aligned} \tag{I–2.5}$$

Substituting these into the equation of state requires $w = -1$ to cancel the Λ terms. This means that Λ provides the negative pressure necessary to cause accelerated expansion [376, Chapter 1].

The evolution of the universe is described by the Friedmann equation, provided

by the 00-component of equation (I–2.3) with an FLRW metric. In conformal coordinates, where Λ has been absorbed,

$$\mathcal{H}^2 = \frac{8\pi G a^2}{3}\varrho(\tau) - \kappa \,. \tag{I–2.6}$$

$\mathcal{H} = a'/a$ is the Hubble function, a is the scale factor first mentioned in equation (I–2.1) and $' \equiv d/d\tau$ is differentiation with respect to conformal time [161, Chapter 18]. It becomes useful to consider the *critical density* [161, Chapter 18], the density required to yield a flat universe,

$$\varrho_\mathrm{c} \equiv \frac{3\mathcal{H}^2}{8\pi G a^2} \,. \tag{I–2.7}$$

Using the critical density, a dimensionless density parameter can be defined as the ratio

$$\Omega \equiv \frac{\varrho(\tau)}{\varrho_\mathrm{c}} \,. \tag{I–2.8}$$

The current total energy density in the universe can be calculated from the measurements of the Hubble function and the density parameter today, $H_0 = \mathcal{H}_0$ and Ω_0, setting the present-day scale factor $a_0 = 1$ [161, Chapter 18]. $\Omega_0 = 1$ corresponds to a currently flat universe. Since the total energy density is formed from the energy densities of radiation, matter and dark energy it becomes useful to define their present density parameters as Ω_r, Ω_m and Ω_Λ respectively, where $\Omega_0 = \Omega_\mathrm{r} + \Omega_\mathrm{m} + \Omega_\Lambda$. The evolution of each species is dictated by the fluid equation derived by the vanishing divergence of the energy-momentum tensor $\nabla_\mu T^\mu{}_\nu = 0$ [264, Appendix B]. In an FLRW universe this is

$$\varrho' = -3\mathcal{H}(\varrho + p) \,. \tag{I–2.9}$$

Solving for an arbitrary equation of state, $p = w\varrho$ gives the evolution of energy density as $\varrho \propto a^{-3(1+w)}$ [161, Chapter 18]. By substituting in the equation of state parameter for each species, the evolution of the universe is given by the evolution of the density of its constituents,

$$\mathcal{H}^2 = \mathcal{H}_0^2\left(\Omega_\mathrm{r} a^{-2} + \Omega_\mathrm{m} a^{-1} + \Omega_\Lambda a^2 - \Omega_\kappa\right) \,, \tag{I–2.10}$$

where $\Omega_\kappa = 1 - \Omega_\mathrm{r} - \Omega_\mathrm{m} - \Omega_\Lambda$ indicates the deviation of the universe from flatness. Equation (I–2.10) describes the background evolution of a ΛCDM universe. Due to the different scaling regimes of each of the constituents, the universe went through different epochs of expansion [109, Chapter 1]. First, radiation was the dominant cause of cosmic expansion, followed by the matter contribution. If the universe were not flat, the curvature term would then come to dominate finally reaching a regime of cosmological constant expansion.

2.2 Cosmic microwave background

2.2.1 CMB temperature

The universe was opaque until approximately 300,000 years after the big bang [109, Chapter 1]. Due to the large amount of energetic radiation when the universe was young, nuclei were instantly ionised [161, Chapter 17]. The mean-free path length of travelling photons was extremely short at this time. It was only as the universe cooled well below the binding energy of electrons and protons, at a temperature of $T \approx 2700\text{K}$, that photons could begin to free-stream [316, Chapter 3]. The free-streaming occurs below the binding energy of electrons and protons because there is such a large proportion of photons to baryons, so the probability of interaction is high [109, Chapter 2]. This surface of last scattering is what is being observed by measurements of the CMB.

Due to collisions with electrons earlier than the surface of last scattering, the CMB photons were in equilibrium as they began to free-stream [109, Chapter 2]. This means that the CMB forms an extremely precise blackbody spectrum [109, Chapter 1]. The temperature of the CMB blackbody evolves proportional to the inverse scale factor, since the energy density of photons is proportional to the fourth power of the temperature and evolves as a^{-4} [109, Chapter 2]. The current temperature is $T_{\text{CMB}} = 2.7255(6)\text{K}$ [131], which is in the microwave band of the electromagnetic spectrum.

2.2.2 Anisotropies

The CMB is a perfect blackbody to one part in ten-thousand [339]. Below this level this anisotropies begin to appear. It is, in fact, these anisotropies which give insight to the interesting physics in the early universe. A favoured paradigm is where they are caused by quantum fluctuations during a period of inflation at the beginning of the universe and form the seeds of all structure seen today [376, Chapter 8]. To see where they come from Einstein's field equations need to be taken beyond zeroth-order. First the components of the metric tensor can be Taylor expanded to

$$g_{\mu\nu} \quad = \quad \bar{g}_{\mu\nu} + \delta g_{\mu\nu} \ , \tag{I–2.11}$$

where $\bar{g}_{\mu\nu}$ is the background components of the metric (equivalent to $g_{\mu\nu}$ as so far used) [109, Chapter 5; 376, Chapter 5]. $\delta g_{\mu\nu}$ describes a perturbation about $\bar{g}_{\mu\nu}$. The Einstein tensor can be recalculated using this expansion which gives, again, a background component $\bar{\mathcal{G}}_{\mu\nu}$ (equivalent to $\mathcal{G}_{\mu\nu}$ before) and the

fluctuation $\delta\mathcal{G}_{\mu\nu}$ [109, Chapter 5; 376, Chapter 5]. In a similar vein, the energy-momentum tensor is expanded to

$$T_{\mu\nu} = (\bar{\varrho} + \delta\varrho + \bar{p} + \delta p)u_\mu u_\nu + (\bar{p} + \delta p)g_{\mu\nu} + \sigma_{\mu\nu} \, . \quad \text{(I–2.12)}$$

It is conventional to denote the perturbed quantities as the density contrast $\delta \equiv \delta\varrho/\bar{\varrho}$ and the pressure contrast $\Pi \equiv \delta p/\bar{\varrho}$ [109, Chapter 5; 376, Chapter 5]. $\sigma_{\mu\nu}$ is a first-order tensor describing the anisotropic stress. The perturbed version of the Einstein field equations can be formed

$$\delta\mathcal{G}_{\mu\nu} = 8\pi G\delta T_{\mu\nu} \, , \quad \text{(I–2.13)}$$

dictating the evolution of the fluctuations until they surpass the perturbative regime [109, Chapter 5].

Since the energy density (and the fluctuations in the energy density) of photons is related to the temperature $\varrho_\gamma \propto T^4$, then the anisotropies in the CMB photons can be found by measuring the temperature across the sky [109, Chapter 8]. The photon temperature anisotropy is defined as

$$\Theta(\hat{n}) = \frac{T(\hat{n}) - T_{\rm CMB}}{T_{\rm CMB}} \, , \quad \text{(I–2.14)}$$

that is, the normalised difference between the temperature in the direction of the unit vector $\hat{n}$ and the CMB temperature $T_{\rm CMB}$ [109, Chapter 8; 376, Chapter 2.6]. The CMB temperature is found by averaging the temperature from all directions on the sky from all positions

$$\begin{aligned} T_{\rm CMB} &= \langle T(\hat{n}) \rangle \, , \\ &= \frac{1}{4\pi}\int d\Omega_{\hat{n}} T(\hat{n}) \, . \end{aligned} \quad \text{(I–2.15)}$$

where $d\Omega_{\hat{n}}$ is the infinitesimal solid angle in a direction $\hat{n}$. $\Theta(\hat{n})$ from equation (I–2.14) is continuous but can be decomposed into a set of discrete spherical harmonics $Y_\ell^m(\hat{n})$,

$$\Theta(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_\ell^m(\hat{n}) \, , \quad \text{(I–2.16)}$$

where $a_{\ell m}$ are an infinite set of complex, constant coefficients, with $\ell = 0 \to \infty$ and $m = -\ell \to \ell$ [109, Chapter 8]. Since the $\Theta(\hat{n})$ are random variables, each of the $a_{\ell m}$ are also random variables [109, Chapter 8]. The spherical harmonics are the analogue of the Fourier series over the surface of a sphere where $\ell = 0$ refers to a function on the whole sphere, $\ell = 1$ over the three axes of the sphere and so on [96]. Higher ℓ therefore corresponds to fluctuations on smaller scales. The $a_{\ell m}$ then describe the amplitude of these fluctuations at every position in the universe [376, Chapter 2.6].

2.3 Power spectrum

The average of the temperature fluctuations over the whole sky vanishes but the variance does not [109, Chapter 8]. The variance is directly related to the angular power spectrum C_ℓ. C_ℓ is a measure of the average amplitude of $a_{\ell m}$ for each ℓ independent of the orientation of the function, given by m [109, Chapter 8; 376, Chapter 2.6]. This is written

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell\ell'}\delta_{mm'}C_\ell \,, \tag{I–2.17}$$

where $\langle\ \rangle$ implies the average over infinite realisations of random processes, i.e. averages over infinitely many different primordial perturbations each with a set of $a_{\ell m}$ [376, Chapter 2.6]. The measured temperature variation $\Theta(\hat{n})$ can be related to the angular power spectrum by taking the ensemble average

$$\langle \Theta(\hat{n})\Theta(\hat{n}') \rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\hat{n}\cdot\hat{n}') \,, \tag{I–2.18}$$

where $P_\ell(\hat{n}\cdot\hat{n}')$ are the Legendre polynomials. To find this, the addition theorem for spherical harmonics [121] has been used, along with equation (I–2.17). The observed power spectrum differs from the theoretical power spectrum of equation (I–2.18) since there is only one available set of $a_{\ell m}$ which can be observed, those of the CMB [109, Chapter 8; 376, Chapter 2.6]. In this case equation (I–2.18) remains the same but the definition of the angular power spectrum changes to

$$\sum_m |a_{\ell m}|^2 = (2\ell+1)C_\ell^{\rm obs} \,. \tag{I–2.19}$$

The expected square of the fractional difference $\Delta C_\ell = (C_\ell^{\rm obs} - C_\ell)/C_\ell$, between the theoretical and observed angular power spectrum reveals a fundamental uncertainty in the measurement of the C_ℓ known as *cosmic variance* [109, Chapter 8; 376, Chapter 2.6] which states that

$$\left\langle \Delta C_\ell^2 \right\rangle = \frac{2}{2l+1} \,. \tag{I–2.20}$$

At low ℓ there is a large uncertainty between the observed $C_\ell^{\rm obs}$ and the theoretical C_ℓ which becomes smaller for large ℓ. It is useful to consider $\Theta(\mathbf{k})$, which is the Fourier transform of $\Theta(\hat{n})$, since the wavenumber k quantifies the size of scales on the CMB [109, Chapter 8]. The temperature anisotropy can be decomposed into an initial condition $\xi(\mathbf{k})$, and a *transfer function* $T_\ell(k)$, which is the solution to the evolution equation of each energy-density species evaluated until

today [109, Chapter 8]. The power spectrum can now be written as

$$\begin{aligned} \langle \Theta(\mathbf{k})\Theta(\mathbf{k}')\rangle &= \int \frac{d^3k}{(2\pi)^3}\int \frac{d^3k'}{(2\pi)^3}\langle \xi(\mathbf{k})\xi(\mathbf{k}')\rangle |T_\ell(k)|^2 \,, \\ &= \int \frac{d^3k}{(2\pi)^3} P_0(k)|T_\ell(k)|^2 \,. \end{aligned} \tag{I–2.21}$$

In the second line $\langle \xi(\mathbf{k})\xi(\mathbf{k}')\rangle = (2\pi)^3 P_0(k)\delta^{(3)}(\mathbf{k}-\mathbf{k}')$ where $P_0(k)$ is the primordial power spectrum which describes any ignorance in initial conditions [109, Chapter 8]. More about the primordial power spectrum will be mentioned in Part I–3. Equation (I–2.21) can be equated to Fourier transform of equation (I–2.18) integrating over all angles on the sky to find the relation between the photon transfer function and the C_ℓ,

$$C_\ell = \frac{2}{\pi}\int dk k^2 P_0(k)|T_\ell(k)|^2 \,. \tag{I–2.22}$$

In the next chapter, more detail about calculating the photon transfer function $T(k)$, will be mentioned.

Chapter 3

Calculating observable effects

3.1 Einstein-Boltzmann equations

To constrain cosmological parameters detailed information about the individual elements which make up the universe and their interactions must be calculated. This is done using the Einstein-Boltzmann equations [109, Chapter 4]. The general form of the Boltzmann equation describes the rate of change of the probability density function of an ensemble of particles due to forces acting on them, their diffusion and any collisions which occur [109, Chapter 4]. It is useful to begin by considering each component and their interaction. CDM provides collisionless, pressureless perturbations to Einstein's field equations [109, Chapter 4; 376, Chapter 6], whilst massless neutrinos are also collisionless but provide pressure since they are a form of radiation [109, Chapter 7]. Massive neutrinos (which can be considered a form of hot or warm dark matter) are again collisionless, but they become non-relativistic once the temperature of the universe drops below their mass and so the momentum dependence cannot be integrated out [109, Chapter 7]. More about this will be mentioned in Part III–2.4.1. Photons and baryons are coupled in the early universe via Compton scattering, and so provide a collision term to the Boltzmann equations [109, Chapter 4; 376, Chapter 6]. Here, "baryons" is a term collectively describing protons and electrons as well as neutral hydrogen, and neutral and ionised helium [109, Chapter 4].

Codes such as `CMBFAST` [327] and `CAMB` [237] treat the Einstein-Boltzmann equations in a particular way. The evolution of the CMB temperature (and polarisation) anisotropies are described using a hierarchy of coupled differential equations [243; 327]. These include a Compton scattering term due to the photon's interaction with baryons as well as the influence from all the other constituents through their coupling to geometry [109, Chapter 4; 243]. By integrating these

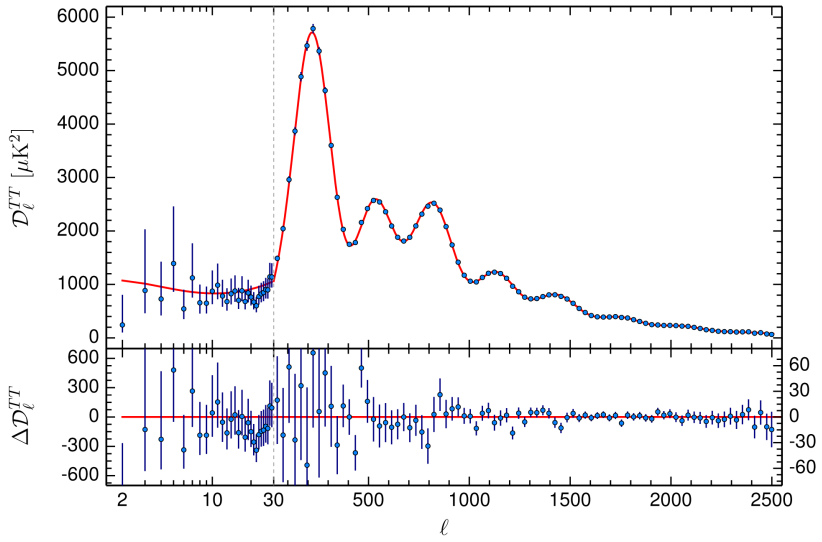


Figure I–3.1: (Top) *Planck*2015 temperature power spectrum with ΛCDM best fit. (Bottom) Residuals with respect to ΛCDM [356].

over the past lightcone, the temperature anisotropies can be written as a geometric term, in the form of Bessel functions, which are independent of cosmology and a source term encoding all the physics [327]. This is very efficient since the Bessel functions can be precomputed on a grid and only the first few photon multipole moments need to be calculated [237]. Since power from the low multipole moments are transferred to the higher multipoles more terms in the expansion need to be kept for accurate computation, but still relatively few are needed compared to calculating each of the multipoles independently [327]. As the power spectrum is smooth then the amplitude of the C_ℓ only needs to be calculated at a sparse number of multipoles, which can then be interpolated between. This is only true for larger $\ell \gtrsim 10$ since the spectrum is discrete at small multipoles [237; 327].

3.2 Power spectrum shape

The values of cosmological parameters can be inferred from the angular power spectrum for the temperature anisotropy. The amplitude of the C_ℓ at a particular ℓ come from the various interactions during the early universe [109, Chapter 8]. The low ℓ anisotropy is dominated by the Sachs-Wolfe effect and, to a lesser extent, the integrated Sachs-Wolfe effect (ISW) [295; 312]. These occur due to the gravitational redshift (or blueshift) of CMB photons due to fluctuations in the gravitational potential at last scattering, or due to time-dependent fluctuations of the gravitational potential between the surface of last scattering and

the present day. The angular power spectrum due to the Sachs-Wolfe effect is proportional to $1/l(l+1)$. For this reason the power spectrum is often plotted as multipole ℓ against $l(l+1)C_\ell$ [68].

The higher ℓ anisotropies come from the effective temperature of the photons and velocity of the radiation-matter fluid at last scattering [335] and further input from the ISW [295].

The main features in the power spectrum from CMB photons is due to the interplay between tightly-coupled photons and baryons moving through a gravitational potential when the universe was still opaque [282; 347] . The baryon-photon fluid is somewhat compressible and at the bottom of a gravitational potential well the photons are in an overdense region causing a maximum in the photon temperature. Pressure in the fluid causes the photons and baryons to rarefact to an underdense region with a minimum in photon-baryon density and minimum in the photon temperature. The fluid collapses again and the oscillations continue until the photons decouple from the baryons at which point the overdensities and underdensities in the temperature are frozen in. These baryon acoustic oscillations (BAO) form peaks in the power spectrum. The first peak is caused by the photon-baryon fluid which has only undergone maximum compression by the time of freeze-out, with the second due to the first maximum rarefaction and so on [109, Chapter 8].

Since baryons are massive they provide drag, slowing the movement of the fluid and thus increasing the frequency of the peaks, but they also increase the gravitational potential [378]. This has the effect of increasing the amplitude of the odd peaks and reducing the even peaks [280].

Doppler shifting of the photon temperature also occurs due to movement in the photon-baryon fluid. At maximum compression and maximum rarefaction the fluid is stationary and thus the Doppler shift has no effect, but inbetween photons temperatures are modified. These are completely out of phase with the BAO and thus have the effect of flattening out the peaks in the power spectrum [109, Chapter 8].

Finally, the photon-baryon fluid is not truly completely coupled since the interaction cross-section is not infinite. This means that there is some diffusion of photons through the baryons which causes a washing out of the peak structure of the power spectrum which becomes more pronounced for higher ℓ [334].

It is through the use of the peak structure of the power spectrum that constraints can be put on the amount of matter, radiation, dark energy and curvature in the universe, amongst many other factors.

3.3 Cosmological parameters

So far, only vague descriptions of how different constituents affect the universe have been discussed. These can be collected into a few cosmological parameters. In ΛCDM in particular there are six parameters which can be tuned to precisely match the observed power spectrum [122]. These parameters are the physical densities of baryons and CDM, the size of the sound horizon at last scattering, the amplitude of primordial scalar perturbations, the spectral tilt of the primordial power spectrum and the optical depth to reionisation which are written $\Omega_{\rm b}h^2$, $\Omega_{\rm c}h^2$, $\Theta_{\rm MC}$, $A_{\rm s}$, $n_{\rm s}$ and $\tau_{\rm R}$ respectively.

The physical densities of baryons and CDM are self-descriptive. They are directly related to the density contrast of equation (I–2.8) and more directly $\Omega_{\rm m} = \Omega_{\rm b} + \Omega_{\rm c}$ of equation (I–2.10). h is a dimensionless number equivalent to $0.01 \times H_0$ km^{-1}s Mpc where H_0 is the Hubble constant. It was originally introduced to quantify the uncertainty in the value of the Hubble constant in historic measurements [317; 365], but can be well constrained now that H_0 is known to within a few percent. Since $\Omega_{\rm b}$ and $\Omega_{\rm c}$ appear directly in the Friedmann equation (I–2.10) they clearly have an effect on the background evolution of the universe. They also affect the temperature anisotropies. As mentioned in the previous chapter, baryons are tightly coupled to photons and result in BAO. If the fraction of baryons in the universe is higher the relative heights of the odd peaks to the even peaks in the power spectrum become larger. Also, due to the greater drag caused by a larger density of baryons, the C_ℓ at higher multipoles become more washed out [187; 335]. The CDM content provides the gravitation potential which initially causes the oscillations to occur [65; 69; 280]. This gravitational potential is also the progenitor of the Sachs-Wolfe effect and the ISW [295; 312]. It is therefore connected with various aspects of the power spectrum, including the small-scale amplitude, the frequency of the BAO peaks and their relative heights [66].

The first peak of the CMB power spectrum describes the fundamental mode of the photon-baryon fluid frozen in at last scattering, as mentioned in the previous chapter. $\Theta_{\rm MC}$ describes the angular size of this *sound horizon*. Although

there is some dependence on the amount of CDM and the value of the Hubble constant, it is mostly determined by the amount of curvature in the universe [95].

The primordial power spectrum $P_0(k)$, first mentioned in equation (I–2.21), quantifies any ignorance in the early universe. It is widely acknowledged that the universe began with a period of exponential growth known as inflation [16; 152]. This solves many issues which arise from observations of the universe, such as: isotropy - the fact that non-causally connected regions look the same; homogeneity - the temperature is the same throughout the universe; and flatness - the curvature of the universe must be highly fine tuned to match observations. It is not in the remit of this Thesis to go into too much detail about inflation. For more information see the review [239]. It is, however, important to note that observable fluctuations in the universe could come from perturbations in the field causing inflation, the inflaton [16]. Variations in the inflaton can be assumed to be Gaussian and (nearly) scale invariant. In the exactly scale invariant case, the primordial power spectrum is described by the Harrison-Zel'dovich spectrum [160; 385]. For deviations from scale invariance, the power spectrum can be quantified with just two parameters, the scalar perturbation and the spectral tilt, A_s and n_s

$$P_0(k) \quad = \quad A_\mathrm{s} k^{n_\mathrm{s}-1} \,. \tag{I–3.1}$$

If $n_\mathrm{s} = 1$ then the spectrum is totally scale invariant and P_0 provides a shift in the amplitude of the CMB power spectrum only. However, if $n_\mathrm{s} \neq 1$ then the tilt in primordial power spectrum influences the scale dependence in the CMB power spectrum.

After recombination, when protons and electrons bound to form neutral hydrogen, photons were able to free stream [109, Chapter 4]. Later, regions of the universe began to become dense again due to gravitational collapse. In these collapsed regions the energy became large enough to ionise the neutral hydrogen, an epoch known as reionisation [109, Chapter 4; 184; 346]. Photons began to scatter off electrons once again, with some fraction of the CMB photons being scattered out of the line-of-sight [67; 282; 348] . The probability that the photons scatter during reionisation is proportional to $\exp(-\tau_\mathrm{R})$ where τ_R is the optical depth to reionisation. τ_R can be found by integrating the electron density and the scattering cross-section from the period of reionisation until today [184]. This causes a dampening in the high multipole tail of the power spectrum. The main effect of reionisation, though, is causing polarisation of the CMB photons. Two types of photon polarisation are seen in measurements of the CMB. These

Parameters	Constraints
$\Omega_{\rm b}h^2$	0.02222 ± 0.00023
$\Omega_{\rm c}h^2$	0.1199 ± 0.0022
$100\Theta_{\rm MC}$	1.04086 ± 0.00048
$\ln\left(10^{10}A_{\rm s}\right)$	3.090 ± 0.039
$n_{\rm s}$	0.9652 ± 0.0062
$\tau_{\rm R}$	0.078 ± 0.019

Table I–3.1: Constraints on the six ΛCDM parameters from the *Planck*2015 temperature and low-ℓ polarisation results [14]. It should be noted that in [19], with the inclusion of the *Planck* low-ℓ EE polarisation likelihood from *Planck* HFI (`lollipop`) analysis $\tau_{\rm R}$ reduces substantially to $\tau_{\rm R} = 0.058 \pm 0.012$.

are the curl-free E-mode and the divergence-free B-mode, named due to their similiarity to the electric and magnetic fields respectively. E-modes are created by photons scattering off electrons, which is the reason why polarisation is expected to arise during reionisation. B-modes on the other hand are created via interactions with gravitational waves [205], although E-modes can leak into the B-mode signal so foreground contamination must be well considered [6]. The discovery of B-mode polarisation from primordial gravitational waves would be an exceptional tool since it could probe into the epoch before last scattering, earlier than the temperature or E-mode polarisation are able to [205; 206; 324; 328; 384].

Although six parameters provides a very simplistic model of cosmology, it is remarkable how well they fit the data. The *Planck*2015+low-ℓ [14] cosmological parameters constraints can be found in Table I–3.1. ΛCDM is often extended by the addition of five other parameters [9]. So far no mention has been made to tensor perturbations, for which new parameters such as the primordial tensor amplitude and tensor spectral tilt are needed. The spectral indices (scalar and tensor) could also be allowed to *run* so the primordial power spectrum is not only not scale invariant but the spectral index is itself scale dependent [53]. The equation of state parameter for dark energy could also deviate from $w = -1$ [340]. A large portion of this Thesis includes extensions to the *vanilla* six parameter ΛCDM. For example, in Part II cosmic string parameters are added to modify the CMB power spectrum to constrain to what extent cosmic strings are present in the universe. Since tensor perturbations are provided by cosmic strings then the ΛCDM tensor parameters are also included. In Part III–2.4.1 massive and sterile neutrino parameters, $\sum m_\nu$, and $m^{\rm eff}_{\rm sterile}$ and $\Delta N_{\rm eff}$ are included to try

and reconcile inconsistent ΛCDM parameters when constrained using the CMB and probes of LSS.

3.4 Markov chain Monte Carlo

With some selected data D, a likelihood function $L(D|\theta, \mathcal{M})$, can be constructed from the theoretical calculations within a given model $\mathcal{M}$ with parameters θ [196]. This is the probability of the data given some parameter values within a model. For example, using RSD (described in more detail in Part III–2.2.2) observations measure three quantities, the Alcock-Paczynski factor, the ratio of the comoving volume distance to the distance to the sound horizon during the drag epoch and the growth rate $\theta = \{F_{\rm AP}(z_{\rm eff}),\, D_{\rm V}(z_{\rm eff})/r_{\rm s}(z_{\rm d}),\, f(z_{\rm eff})\sigma_8(z_{\rm eff})\}$ [143]. The correlation between these parameters can be collected in a covariance matrix $C_{\rm RSD}$. A likelihood can then be produced by considering the difference between the data and the theoretical model, say ΛCDM, prediction

$$L(\mathrm{RSD}|\ \theta, \Lambda\mathrm{CDM}) \quad \propto \quad \exp\left[-\frac{(\theta_{\rm obs}-\theta_{\rm pred})C^{-1}_{\rm RSD}(\theta_{\rm obs}-\theta_{\rm pred})}{2}\right]. \quad \text{(I–3.2)}$$

$\theta_{\rm obs}$ and $\theta_{\rm pred}$ are the values of the three RSD parameters from observations and calculated from a set of cosmological parameters in the theoretical model respectively. $C^{-1}_{\rm RSD}$ is the inverse of the observed RSD covariance matrix.

To be able to learn about the probability of parameters then the posterior distribution needs to be calculated [196]. This can be found using Bayes' theorem

$$P(\theta|D, \mathcal{M}) \quad = \quad \frac{L(D|\theta, \mathcal{M})p(\theta)}{p(D)}, \quad \text{(I–3.3)}$$

where $p(\theta)$ is the prior on the values of θ and

$$p(D) \quad = \quad \int d\theta L(D|\theta, \mathcal{M})p(\theta) \quad \text{(I–3.4)}$$

is the evidence. The whole distribution up to a normalisation constant could, in theory, be constructed on a grid [178, Chapter 3]. For high dimensional problems such as in cosmology, this is prohibitively expensive. A more efficient way to understand the probability of cosmological parameters given the data is to draw samples from the posterior distribution. Statistics can then be obtained from these samples. There are various methods to obtain samples from a posterior distribution, but one of the most efficient for cosmology is the use of Markov chain Monte Carlo (MCMC) [178, Chapter 3; 234].

A Markov chain is constructed of vectors of parameter samples where each subsequent vector is related to only the previous vector and no others [329, pg. 2]. A good example of a Markovian process is a *random walk* where each step depends on the current position, but not any of the previous positions [132, pp. 373-374]. Markov chains can be used to sample from a probability distribution by dictating that the probability of arriving at a point θ_i is proportional to the probability distribution at that point $P(\theta_i)$. The probability at the next point in the chain θ_{i+1}, must be equivalent to the probability of transferring to that point from θ_i ([132, pp. 373-374])

$$P(\theta_{i+1}) \quad = \quad \int d\theta_i P(\theta_i) T(\theta_i, \theta_{i+1}). \tag{I–3.5}$$

$T(\theta_i, \theta_{i+1})$ is a transfer probability which must be defined such that the probability of transferring *here* and being *there* is the same as being *here* and transferring *there*

$$P(\theta_{i+1}) T(\theta_{i+1}, \theta_i) \quad = \quad P(\theta_i) T(\theta_i, \theta_{i+1}). \tag{I–3.6}$$

This is known as *detailed balance* [178, Chapter 3]. The samples obtained using detailed balance will probe the underlying distribution, but neighbouring samples will be correlated. For this reason many steps need to be taken between points to ensure mostly uncorrelated samples.

The transfer function used for parameter constraints throughout this Thesis comes from the Metropolis-Hastings algorithm [163; 262]. The transfer probability is

$$T(\theta_i, \theta_{i+1}) \quad = \quad \alpha(\theta_i, \theta_{i+1}) q(\theta_i, \theta_{i+1}), \tag{I–3.7}$$

where $q(\theta_i, \theta_{i+1})$ is a proposal density distribution and

$$\alpha(\theta_i, \theta_{i+1}) \quad = \quad \text{Min}\left[1, \frac{P(\theta_{i+1}) q(\theta_{i+1}, \theta_i)}{P(\theta_i) q(\theta_i, \theta_{i+1})}\right], \tag{I–3.8}$$

is the probability that a new point is accepted. It is usually chosen such that the proposal density is symmetric $q(\theta_{i+1}, \theta_i) = q(\theta_i, \theta_{i+1})$ [178, Chapter 3]. This means that if the probability is low at an initial point θ_i and the probability is higher at the next θ_{i+1} then $\alpha = 1$ and so the new proposed point is accepted. On the other hand, if the probability at the next point is lower than at the current point then there is some probability that the next point will be rejected and another "next step" must be taken to a different point. If a point is rejected the step is effectively taken to the same point. This can be added to the samples as the next step, but clearly is correlated with the current point [178, Chapter 3].

It is important that the proposal density approaches the true distribution which is being probed to cover the whole parameter space with samples [304]. If the proposal density is too narrow, once the peak of the distribution is reached there is very little probability that any samples from the edges of the distribution will be obtained. Likewise, if the proposal density is too broad, the steps will often stray away from the peak of the distribution and not give a representative sample of points from the distribution since the acceptance probability will be low. Fortunately the proposal density can be tuned during the early stages of running the Markov chain [178, Chapter 3]. This early stage is known as *burn-in* and is a period where the points are randomly probing the distribution but have not reached the maximum and so are not drawing representative samples. Once the maximum is reached and the proposal density is tuned using the covariance matrix from the early samples, these burn-in samples are discarded and the Markov chain begins in truth. Since each adjacent sample is correlated it is important to obtain very large numbers of samples to make sure each is independent. With a perfect proposal density it takes n steps to cover the distribution in each parameter direction if the steps are of size $1/\sqrt{n}$. This means it takes $\mathcal{O}(n)$ steps to cover the distribution in each direction and hence each $\sim n^{\text{th}}$ step is as close to uncorrelated as possible [178, Chapter 3].

The collection of samples forms an efficient high dimensional representation of the posterior distribution $P(\theta|d,\mathcal{M})$ from which statistics can be calculated. When quantifying single parameters, each of the other parameters must be integrated out, a technique known as marginalisation [358, pp. 32–33]. For example, say $\theta = \{\alpha,\beta,\gamma,\delta\}$ is a vector of parameters then the posterior distribution of α is

$$P(\alpha|D,\mathcal{M}) \quad = \quad \int d\beta\ d\gamma\ d\delta\ P(\alpha,\beta,\gamma,\delta|D,\mathcal{M}). \tag{I–3.9}$$

It is not trivial to go from $P(\alpha|D,\mathcal{M})$ back to $P(\alpha,\beta,\gamma,\delta|D,\mathcal{M})$ and so care should be taken when interpreting the marginal distribution.

Another useful property of the samples obtained via MCMC is that they can be importance sampled. Importance sampling is an efficient way of getting samples from a new distribution which is very similar to the posterior the samples were drawn from [178, Chapter 3]. For example, if the choice of prior used whilst obtaining the samples is changed, perhaps due to the results of new data, then the likelihood of the samples in the chains can be re-weighted under the influence of the updated prior. The weights are derived from the ratio of the

probability of obtaining a sample in the new distribution to the probability of obtaining the same sample in the original distribution [178, Chapter 3]. New statistics, such as the expected value of a parameter, can then be easily found by summing the value of that parameter for every sample weighted by the weight of each corresponding sample (normalised by the sum of the weights). Again, care needs to be taken here. If the new distribution is distinctly different from the original, then the weights will be close to zero for much of the parameter range. Hence, most of the samples will be lost and the expected parameters from the new distribution may be skewed from their values when considering samples drawn directly from the new distribution.

With well defined posterior distributions and known priors the plausibility of two models can be tested. The Bayesian evidence is one way to compare two models $\mathcal{M}_1$ and $\mathcal{M}_2$. This takes into account relative sizes of the model spaces allowed, penalising complicated models with large number of parameters and a significant amount of freedom and favouring simpler models. Care needs to be taken when using the evidence since it is sensitive to the choice of priors (more is mentioned about this in Part III–2.4.1). Typically when a likelihood approach prefers the inclusion of the parameter at $> 3\sigma$, the use of Bayesian evidence will come to the same conclusion. The Bayes factor quantifies the relative plausibility of two models with the same *a priori* probability

$$\Theta_{\mathrm{B}} = \frac{P(\mathcal{M}_1|D)P(\mathcal{M}_2)}{P(\mathcal{M}_2|D)P(\mathcal{M}_1)}, \tag{I–3.10}$$

where $P(\mathcal{M}|D)$ is the conditional probability of a model being correct given the data, D, and $P(\mathcal{M})$ is the probability of the model being correct [195]. The model probabilities are usually normalised such that $P(\mathcal{M}_2)/P(\mathcal{M}_1) = 1$. When $\mathcal{M}_1 \subset \mathcal{M}_2$ then the Savage-Dickey density ratio can be used to simplify the Bayes factor (see e.g. [367] for details)

$$\Theta_{\mathrm{B}} = \left.\frac{P(\psi|D,\mathcal{M}_2)}{P(\psi|\mathcal{M}_2)}\right|_{\psi=\psi_1} \tag{I–3.11}$$

where ψ are the additional parameters in the extended model and ψ_1 are their fiducial values in the nested model. Therefore, in order to calculate Θ_{B} only the parameter posterior likelihood is needed for the extended model and the probability defined by the prior at the value the parameter would have in the base model.

Part II

Cosmic strings

Cosmic strings are line-like concentrations of energy that can arise as topological defects in theories of the early Universe [86; 91; 172; 215; 216; 372]. In particular, they form naturally in models of hybrid inflation [48; 75; 93; 117; 118; 200; 240] in which the inflationary phase ends with a second-order phase-transition [75; 90; 194; 318]. Although they were originally considered as an alternative candidate for providing the seeds for structure formation in the Universe [71; 85; 217; 283], it is now understood that they cannot give rise to the observed BAO in the power spectrum [22; 23; 32; 47; 89], but can play a subdominant role. There are a wide range of potential observational signatures of cosmic strings, including wakes in the density of matter perturbations, gravitational waves from loop radiation and line-like discontinuities in the CMB temperature anisotropy via the Kaiser-Stebbins effect [74; 148; 203; 372, Chapter 11]. Thus, strings provide a powerful tool for testing theories of the early Universe. Observations have strongly constrained the contribution of cosmic strings to the total CMB anisotropy [13; 23; 46; 120; 290; 333; 381]. Current data place a 2σ upper bound on the string tension of $G\mu < 1.3 \times 10^{-7}$ for Nambu-Goto strings [14] or $G\mu < 2.7 \times 10^{-7}$ for Abelian-Higgs strings [241], which corresponds to $\sim$1% of the total temperature anisotropy at $\ell = 10$. G is the gravitational constant, μ is the tension of the string. Although this may seem insignificant, there is still constraining power left in the data since strings generate specific signatures in the primordial B-mode polarisation spectrum [59; 266; 289; 290; 325; 326; 362], which can now be analysed with the *Planck*2015 polarisation [18] and joint BICEP2 data [6].

Going beyond the simplest cosmic string models, complex networks of multiple types of interacting superstrings, each with a different tension, can also be considered. Notably, interacting networks of fundamental F-strings, one dimensional D-branes (D-strings) and bound (FD) states between F- and D-strings, collectively referred to as cosmic superstrings, arise naturally in string theoretic inflation [75; 119; 360]. These networks are notably different to their simpler, single-type string counterparts since the different string types have intercommutation probabilities that are not necessarily unity [34; 35; 157; 192; 193; 198; 360]. The interactions among different string types are also much more complex, as colliding strings can zip together or unzip, producing heavier or lighter FD-string states carrying different charges. These features affect CMB signatures allowing us to obtain constraints on string theory parameters such as the string coupling $g_{\rm s}$ and the fundamental string tension $\mu_{\rm F}$ [36; 293].

In Part II–1 a brief overview of the formation of cosmic strings via sponta-

neous symmetry breaking is presented. Along with this, the scaling solutions of networks of cosmic strings and their observational signatures are shown. In Part II–2 the *Planck*2015 public data [18] is used to perform the first full MCMC analysis of ΛCDM models with cosmic string or superstring networks. For "ordinary" cosmic string networks the unconnected segment model (USM) framework is used and the analytic method introduced in [37] is used for fast computation of the string unequal-time correlator (UETC). This is used as a source to compute CMB anisotropies and hence obtain joint constraints on ΛCDM and the string network parameters, including the tension $G\mu$, the loop chopping efficiency $c_{\rm r}$ and the wiggliness parameter α. In the case of cosmic superstring networks the method is extended to deal with multiple network components. The UETC approach is efficient, meaning the superstring spectrum can be computed in much less time than previous codes, obtaining joint constraints on the fundamental string tension $G\mu_{\rm F}$, the string coupling constant $g_{\rm s}$, the self-interaction coefficient $c_{\rm s}$, and the parameter w of [293], quantifying the volume of compact extra dimensions.

Chapter 1

Cosmic strings

1.1 Spontaneous symmetry breaking

Cosmic strings are topological defects arising from the breaking of axial symmetries [216]. If in fieldspace, there is a non-simply connected vacuum manifold, then non-contractible loops generically form. For example, take the local Abelian scalar field (Higgs) theory ([270])

$$\mathcal{L} = -\frac{1}{2}\bar{\nabla}_\mu\bar{\varphi}\nabla^\mu\varphi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(\varphi), \tag{II–1.1}$$

where $\varphi(x)$ is a complex scalar field ($\bar{\varphi}$ is the complex conjugate), ∇_μ is a covariant derivative and $F_{\mu\nu}$ is the field strength tensor of the Abelian gauge field A_μ,

$$\nabla_\mu\varphi = (\partial_\mu + ieA_\mu)\varphi, \tag{II–1.2}$$

where e is the gauge coupling constant and

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{II–1.3}$$

The bare potential is quartic in the fields ([171])

$$V_{\text{bare}}(|\varphi|) = \frac{\lambda}{4}(|\varphi|^2 - \eta^2)^2, \tag{II–1.4}$$

where η is the symmetry breaking scale and λ is the coupling to this term. High temperature corrections need to be included and are of the form

$$V(|\varphi|) = V_{\text{bare}}(|\varphi|) + CT^2|\varphi|^2 + \cdots \tag{II–1.5}$$

where the precise details of C are unimportant here, but depend on one-loop corrections [55; 111; 377]. A critical temperature can then be defined by minimising the potential and equating the φ terms to get

$$T_C^2 = \frac{\lambda\eta^2}{2C}. \tag{II–1.6}$$

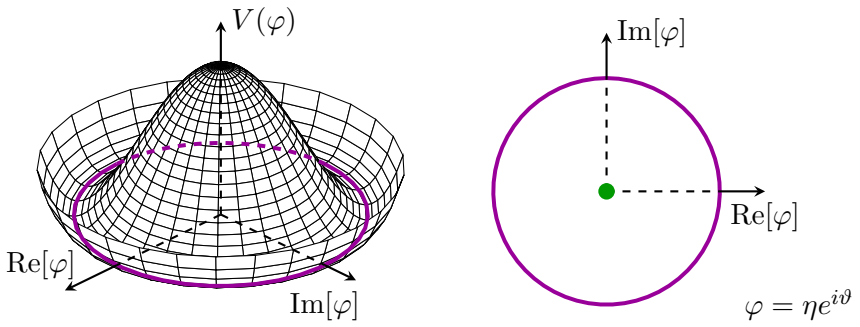


Figure II–1.1: Manifold of allowed field values once the Abelian-Higgs symmetry is spontaneously broken and settled at $T = 0$.

With this, the potential can be rewritten as

$$V(|\varphi|) = \frac{\lambda}{4}\left[|\varphi|^4 + \eta^4 - 2\left(1 - \frac{T^2}{T_C^2}\right)|\varphi|^2\eta^2\right]. \tag{II–1.7}$$

When $T_C > T$ then the positive quadratic term dominates and there is a minimum at $\varphi = 0$. Below T_C the symmetry is spontaneously broken such that the fields obtain vacuum expectation values with magnitudes ([216])

$$|\varphi|^2 = \eta^2\left[1 - \frac{T^2}{T_C^2}\right]. \tag{II–1.8}$$

The manifold of allowed field states is a circle, $\mathcal{M} = U(1)$, i.e. the homotopy group $\pi_1(\mathcal{M}) = \mathbb{Z}$ is non-trivial [216; 265, Chapter 4.6.3; 372, Chapter 3]. At zero temperature this simply reduces to $|\varphi|^2 = \eta^2$. The allowed values of φ can be seen in figure II–1.1. The field values can be represented by a one-parameter family given by

$$\varphi = \eta\, e^{i\vartheta}, \tag{II–1.9}$$

where $0 \leq \vartheta < 2\pi$ [372, Chapter 4]. As the early universe cools below T_C then patches Hubble distances apart settle to different values of the phase ϑ because thermal fluctuations in φ are large [215; 216]. Due to $\bar{\nabla}_\mu\bar{\varphi}\nabla^\mu\varphi$, it is energetically favourable for the regions between these patches to have a smooth transition between values of φ [215]. Since φ must be single valued then any closed loop in space must undergo an integer number of n rotations such that $\Delta\vartheta = 2\pi n$. When $n = 0$ then the closed loop can be deformed to a point and no string exists, but when $n \neq 0$ the loop contains a region where φ is undefined within the manifold of allowed field values [216; 265, Chapter 4.6.3; 372, Chapter 3]. This is the cosmic string. In three dimensional space, this string must be infinitely long, or closed [265, Chapter 4.6.3]. If this were not true then the closed path around the string would be able to deform around the end of the string and close to a

point. Not every edge of a domain forms a cosmic string. When φ in each cell is uncorrelated and varies along the shortest path in $\mathcal{M}$ between the two cells then one in four curves forms a cosmic string [265, Chapter 4.6.3]. An example of cosmic strings forming between domains of different ϑ values can be seen in figure II–1.2. The correlation length ξ can be defined as the distance within which regions of space have correlated ϑ [372, Chapter 2]. As the universe cools, the symmetry is spontaneously broken. Just below the critical temperature, the fluctuations of the field φ are large enough to exceed the potential barrier. Cooling continues, the effective mass of the field becomes finite and the scale of the fluctuations define the correlation length of the string ([215])

$$\xi \sim \frac{1}{\lambda\eta}. \tag{II–1.10}$$

The correlation length is frozen in because the fluctuations in the field are no longer large enough to scale the potential barrier [387]. The temperature where this occurs is known as the Ginzburg temperature and is $\sim \lambda T_C$ lower than the critical temperature and so is $\mathcal{O}(\eta)$ [215]. The horizon size at this temperature is $\mathcal{H} \sim \sqrt{G}\eta^2$ where G is Newton's constant. This shows that the correlation length at the time of formation is proportional to the size of the horizon ([372, Chapter 9])

$$\xi\mathcal{H} \sim \sqrt{G}\,\frac{\eta}{\lambda}. \tag{II–1.11}$$

The strings which form via this mechanism have the structure of a random walk. Simulations of cosmic string formation show that 65–80% of strings are long and the rest are closed loops [31; 363]. The correlation length evolves throughout time, but cannot exceed a Hubble distance. This is because correlations cannot form at faster than the speed of light [216].

1.2 Size and tension of a string

By imposing $\partial_\mu A^\mu = 0$ on the gauge field from equation (II–1.2) integrating a closed path around the string reveals a quantised magnetic flux flowing along the string ([3])

$$\begin{aligned} \Phi_{\mathrm{B}} &= \oint dl \cdot \mathbf{A} \\ &= \frac{2\pi n}{e}. \end{aligned} \tag{II–1.12}$$

Due to this localisation there is vanishing divergence of the scalar field and gauge boson away from the core of the string [270]. The width of the string can then be found from the Compton wavelength of these fields, i.e. the inverse

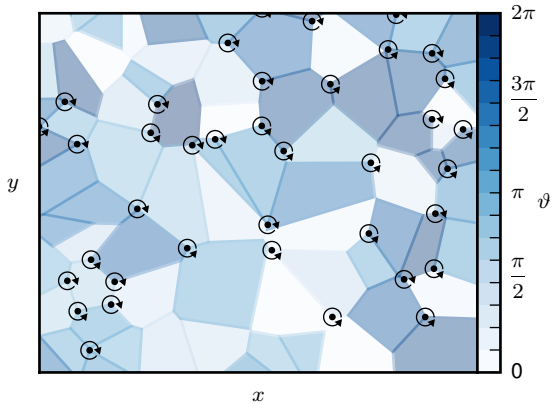


Figure II–1.2: Cosmic string formation when different regions of the universe spontaneously break into different field values.

of the mass of the fields. The Abelian-Higgs' mass is read from the quadratic term from equation (II–1.7) as $m_\varphi = \sqrt{\lambda}\eta$, whilst the gauge boson's mass is $m_A = e\eta$ [372, Chapter 3]. The energy density (tension) of the string is then found by integrating the 00-component of the energy momentum of the string. For a string aligned along the z axis, the $x - y$ plane is integrated

$$\mu \quad = \quad \int dxdy \; T_0{}^0, \tag{II–1.13}$$

where the energy-momentum tensor $T_{\mu\nu}$ is obtained by varying the action with Lagrangian from equation (II–1.1) with respect to the metric [264, Chapter 21]. When the mass of the gauge boson is larger than the mass of the scalar field, the string tension is approximately $\mu \approx \eta^2$. The string is described by a region where $\varphi \to 0$, inside of which is a tube of magnetic flux where $\Phi_\mathrm{B} = 2\pi n/e$. Any small scale structure on the string gets stretched with cosmological expansion, and as a result reduces in amplitude [369]. This leads to long straight strings with their radius of curvature much greater than their width [372, Chapter 6]. Since the width of a cosmic string is much less than the correlation length, cosmic strings can be well modelled by the Nambu-Goto action. This is the description of a one dimensional string with zero with. It is strings of this form that will be considered in Part II–2.

1.3 String evolution

Networks of strings undergo three main processes which dictate their evolution. In fact, these processes interact in such a way that a network of cosmic strings

will evolve towards a scaling regime, where the characteristic length scale of the network is proportional to the size of the horizon [24]. Firstly, the length of any string increases (gets stretched) as the universe expands. If the characteristic length scale of a network of Brownian random walks is L then the stretching dictates this to evolve as

$$L(\tau) = \frac{\mathcal{H}_0}{\mathcal{H}} L(0), \tag{II–1.14}$$

where $\mathcal{H}$ is the comoving Hubble function, $\mathcal{H}_0$ is the value of the Hubble constant today and $L(0)$ is the length scale of the network at an initial conformal time $\tau = 0$ [372, Chapter 9]. The energy for such a network is ([24; 372, Chapter 9])

$$E = \frac{\mu V}{L^2}. \tag{II–1.15}$$

Loops form when a string intersects itself or intersects with another string in the network. Since a string is expected to encounter another string within a distance L in a volume L^3 for a random walk, then a rate of energy density loss can be described by ([372, Chapter 9])

$$E' \approx \left(\mathcal{H} - \frac{a}{L}\right) E. \tag{II–1.16}$$

Defining $\xi\tau = L/a$ then

$$\frac{\xi'}{\xi} = -\frac{3\mathcal{H}}{2} - \frac{1}{2\tau}\left[2 - \frac{1}{\xi}\right]. \tag{II–1.17}$$

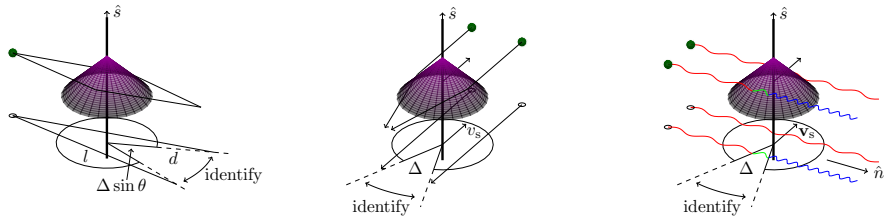
During matter and radiation domination $\mathcal{H} = 2\tau^{-1}$ and $\mathcal{H} = \tau^{-1}$ respectively [109, Chapter 2]. This means that

$$\left(\frac{\xi'}{\xi}\right)_{\text{matter}} = -\frac{1}{2\tau}\left[8 - \frac{1}{\xi}\right], \tag{II–1.18}$$

and

$$\left(\frac{\xi'}{\xi}\right)_{\text{radiation}} = -\frac{1}{2\tau}\left[5 - \frac{1}{\xi}\right], \tag{II–1.19}$$

for matter and radiation. There are regimes, when $\xi_{\text{matter}} = 1/8$ and $\xi_{\text{radiation}} = 1/5$ where the correlation length of the network is fixed with respect to the size of the causal horizon [24; 87]. The correlation length is small when there is a high density of strings [372, Chapter 9] which means that $\xi'/\xi > 0$ and so loop chopping becomes very efficient and the correlation length increases as the density reduces. Likewise, if the correlation lengths are long, when densities are low, $\xi'/\xi < 0$ and loop chopping becomes less efficient and the density increases, reducing the correlation lengths [372, Chapter 9]. The final process which is important for the evolution of string networks is the loss of energy by gravitational radiation at kinks and from small loops [371]. This means that the density of strings will reduce over time.



(a) Doubling of observed images from an object.

(b) Formation of wakes in matter.

(c) Doppler shifting of photons (Kaiser-Stebbins effect).

Figure II–1.3: Observational signatures due to the conic spacetime about a cosmic string

1.4 Observational signatures from cosmic strings

There are several observable effects that cosmic strings are able to provide. These arise from the effect of the cosmic string's energy density on spacetime. To linear order the perturbation of the metric by the string is given (in cylindrical coordinates) as ([372, Chapter 10])

$$ds^2 = -dt^2 + dz^2 - dr'^2 - r'^2 d\theta'^2, \tag{II–1.20}$$

where r' is a radial coordinate about the string related to the usual radial coordinate $\left(r = \sqrt{x^2 + y^2}\right)$ by

$$r'^2 = \frac{1 - 8G\mu \ln(r/r_0)}{1 - 8G\mu} r^2, \tag{II–1.21}$$

with an integration constant r_0 which can be set to the radius of the string and $r > r_0$ [370]. The logarithm arises by solving the perturbed Einstein equations [370; 372, Chapter 6]. θ' from equation (II–1.20) is

$$\theta' = (1 - 4G\mu)\theta, \tag{II–1.22}$$

where θ is the usual polar coordinate. Equation (II–1.20) describes a conical spacetime since $0 \leq \theta' < 2\pi(1 - 4G\mu)$ and so there is an azimuthal deficit of $\Delta = 8\pi G\mu$ [148; 203; 372, Chapter 10]. Due to the conical metric around a cosmic string then both matter and radiation are effected.

The most simple observational example is the bending of the paths of light which pass perpendicular to the string. An example of this mechanism is plotted in figure II–1.3a. Objects on the far side of a cosmic string from an observer will be seen as images to the left and to the right of a string aligned along the z

axis. More generally the missing angle is $\Delta \sin\theta$ and the angular separation of the images is ([372, Chapter 10])

$$\varsigma = \frac{l\Delta \sin\theta}{l+d}, \tag{II–1.23}$$

where l is the distance of the object to the string, d is the distance of the string to the observer and θ is the angle that the cosmic string makes with the plane containing the object and the observer. These long straight-line doubling of images could be visible in weak lensing surveys. For example, thousands of weakly lensed objects would be expected from just one Hubble length cosmic string with $G\mu \approx 10^{-6}$ by LSST (Large Synoptic Survey Telescope) [189], but the signal would mostly be dominated by weak lensing from LSS. Perhaps the use of data mining techniques for large field surveys will be able to access hidden correlations left by cosmic strings in future surveys.

Wakes in matter can also form as a string passes through dust. A velocity perturbation is provided to the matter with magnitude

$$\delta v = v_{\rm s}\gamma\Delta \tag{II–1.24}$$

where $\gamma = (1-v_{\rm s}^2)^{-1/2}$ and $v_{\rm s}$ is the velocity perpendicular to the string [342; 372, Chapter 11]. An example plot is shown in figure II–1.3b. These straight line structures would be observable if the string tension were large enough. For example, future CMB polarisation experiments should be able to detect the signatures of these wakes if the tension is $G\mu \sim 3\times 10^{-7}$ [99].

Most important for this Thesis is the (Gott-)Kaiser-Stebbins effect [148; 203]. An example of this effect is plotted in figure II–1.3c. Discontinuous line features in the frequency of CMB photons, passing on either side of a moving string which is perpendicular to the line of sight, form due to Doppler shifting

$$\Theta^{\rm string} = \gamma\Delta\ \hat{n}\cdot(\mathbf{v}_{\rm s}\times\hat{s}), \tag{II–1.25}$$

where $\Theta^{\rm string}$ is the fractional change in temperature over the step, $\mathbf{v}_{\rm s}$ is the velocity of the string (with magnitude $v_{\rm s}$), $\hat{s}$ is the unit vector along the string and $\hat{n}$ is the unit vector along the line of sight [372, Chapter 11]. The conclusion of this effect is that, photons in front of the moving string remain at the same temperature whilst the deficit angle in the conic spacetime causes photons behind the moving cosmic string to blue-shift [203]. This means that hot to cold line discontinuities in the CMB temperature map would arise.

As already mentioned, cosmic strings emit gravitational radiation either at kinks

in long strings [210] or from the dissipation of loops [74; 364]. Whilst the precise details are not needed here, it is useful to note that strings can provide a stochastic background of gravitational waves which can potentially be detected by future gravitational wave experiments.

1.5 Other topological defects

So far, only cosmic strings have been discussed. Other topological defects can arise from the breaking of different symmetries [216]. The most simple to consider is the breaking of an $O(1)$ symmetry to S^0, i.e. two discrete values. Using the $O(1)$ (real) analogue of equation (II–1.5) at $T = 0$ the field can take values of $\varphi = \pm\eta$ [216; 265, Chapter 4.6.3]. Patches of space which have spontaneously broken to different field values cannot smoothly deform since the manifold of allowed field values is disconnected [215; 216]. A domain wall forms at the boundary between the patches, the region where the disallowed field value $\varphi = 0$ occurs. The energy per unit surface area can be worked out similarly to the energy per unit length of a string. If they are present, there should be at least one domain wall per Hubble time, with a mass $M_{\rm wall} \sim 10^{65}\lambda^{1/2}(\eta/100{\rm GeV})^3$ [265, Chapter 4.6.3]. This is exceptionally large for any non-fine-tuned values of λ and η, which would cause obvious (dominant) fluctuations in the CMB [265, Chapter 4.6.3]. Such fluctuations have not been seen, suggesting that domain walls are not present in the universe [13].

Monopoles can form when the manifold of allowed states forms non-contractible two-surfaces [181; 292]. In this case $\pi_2(\mathcal{M})$ needs to be non-trivial. An example where monopoles arise is the breaking of an $SO(3)$ symmetry. Since the two-sphere, $\mathcal{M} = S^2$, is invariant under $SO(3)$ then it defines the manifold of allowed field values once the symmetry is broken. One particularly important feature of monopoles is that they have a magnetic flux, but this is trapped at a localised point, in a similar way to charged particles sourcing an electric field [106; 181]. Bounds are generally given on the monopole flux $F_{\rm M} = n_{\rm M}v$ where v is the average magnitude of the velocity of monopoles and $n_{\rm M}$ is the number density of monopoles. The Parker bound states that ([277])

$$F_{\rm M} \leq 10^{-15}{\rm cm}^{-2}{\rm s}^{-1}, \tag{II–1.26}$$

which is found by considering the magnetic energy dissipation of galaxies. As the galactic magnetic field accelerates monopoles out of the galactic halo, the field is dissipated. Measuring the current magnetic field of galaxies therefore provides a constraint on the monopole flux [372, Chapter 14]. The strongest

bounds are found by calculating the number of monopoles captured by neutron stars giving ([105; 133; 224])

$$F_{\rm M} \lesssim 10^{-20}{\rm cm}^{-2}{\rm s}^{-1}. \quad \text{(II–1.27)}$$

Since the flux is related to the number density of monopoles, these bounds can be related directly to the energy density of monopoles, which scales with Higgs correlation length [372, Chapter 14]. Monopoles annihilate with anti-monopoles, but the rate of annihilation is lower than observational bounds allow [130]. Not observing as many monopoles as predicted was one of the initial reasons for the introduction of inflation into cosmological evolution [123; 153]. A period of inflation would dilute the number density of monopoles considerably, allowing them to exist but not break observational bounds [152].

Finally, *textures* arise when $\pi_3(\mathcal{M})$ is non-trivial [336; 372, Chapter 15]. In this case, the scalar fields are always in the vacuum manifold but energy remains from the gradient of the field. Knots in the manifold form when regions of space have different field values. These unwind at relativistic speeds which give the fields enough energy to get over potential barriers in the manifold. The energy is then dissipated leaving regions at the same field value. The gravitational fields of textures distort the CMB temperature anisotropies revealing a spectrum like the Harrison-Zel'dovich spectrum, which is promising [372, Chapter 15]. Unfortunately, due to the nature of the high energy unwindings in the early universe, there would be a large non-Gaussian signature which is tightly constrained. It is usual to quote bounds on $G\mu$, as it is for strings, but the string tension is associated with the symmetry breaking scale for the texture ([13])

$$\mu = 2\pi\eta^2. \quad \text{(II–1.28)}$$

From temperature maps alone the constraints from *Planck* on textures is

$$G\mu < 1.06 \times 10^{-6} \quad \text{(II–1.29)}$$

but the sensitivity of non-Gaussian probes of cosmology could reduce this. Most presented results are for global cosmic strings rather than textures [13].

Chapter 2

Cosmic string spectra

2.1 Unequal-time correlator

Unlike passive inflationary perturbations which are set as initial conditions, metric perturbations from cosmic string networks are actively sourced at all times. To compute string spectra the components of the string network's energy momentum tensor must be used as sources in the linearised Einstein-Boltzmann equations. The relevant quantity to calculate is the UETC, whose dominant eigenmodes, found by diagonalising, can be used as source functions, each individual mode being coherent [283]. The UETC

$$\langle \Theta_{\mu\nu}(\mathbf{k},\tau)\Theta^{*}_{\alpha\beta}(\mathbf{k},\tau')\rangle \quad \equiv \quad \mathcal{C}_{\mu\nu,\alpha\beta}(\mathbf{k},\tau,\tau') \tag{II–2.1}$$

determines all the two-point correlation functions such as the CMB temperature C_ℓ and matter power spectra $P(k)$, defined as in [72]. $\Theta_{\mu\nu}(\mathbf{k},\tau)$ is the string energy-momentum tensor defined below.

2.1.1 String energy-momentum tensor

Nambu-Goto strings are one-dimensional defects in the zero-width limit. They provide a good description for long cosmic strings, whose correlation length is many orders of magnitude larger than their width, at least away from string intersections. A string moving in spacetime spans a two-dimensional surface, the worldsheet $x^\mu(\sigma^a)$, where the indices $\mu = 0,1,2,3$ label spacetime coordinates and $a = 0,1$ are the indices of coordinates on the worldsheet [175; 268]. The worldsheet action is reparametrisation invariant and a gauge can be chosen by imposing two conditions on the spacetime coordinates x^μ as functions of σ^a. In an FLRW background, a useful choice of gauge is such that $\sigma^0 = \tau$, the conformal time, and $\mathbf{x}' \cdot \dot{\mathbf{x}} = 0$, where $\dot{} \equiv \partial/\partial\tau$ and $' \equiv \partial/\partial\sigma$, relabelling σ^1, which in this gauge is a spacelike worldsheet coordinate, as σ. In this gauge the Nambu-Goto

string energy-momentum tensor is

$$\begin{aligned}\Theta^{\mu\nu}(y) &= \frac{1}{\sqrt{-g}}\int d\tau d\sigma\left[U\sqrt{-\frac{\mathbf{x}'^2}{\dot{\mathbf{x}}^2}}\dot{x}^\mu\dot{x}^\nu\right. \\ &\quad \left.-T\sqrt{-\frac{\dot{\mathbf{x}}^2}{\mathbf{x}'^2}}x'^\mu x'^\nu\right]\delta^{(4)}(y-x(\tau,\sigma)).\end{aligned} \tag{II–2.2}$$

Here, U is the string energy per unit length and T is the string tension. For Nambu-Goto strings on arbitrarily small scales, Lorentz invariance requires that $T = U = \mu$. However, if the string is coarse-grained, then the integrated effect of small-scale structure is to make the effective tension smaller than the energy density [79; 180]. The effect of small-scale wiggles on the string can then be included via a "string wiggliness" parameter α, such that $U = \alpha\mu$ and $T = \mu/\alpha$ satisfying $UT = \mu^2$.

The Fourier transform of the 00-component of the energy-momentum tensor of a representative string segment in a network is

$$\Theta_{00}(\tau,\mathbf{k},\chi) = \frac{\mu\alpha}{\sqrt{1-v^2}}\frac{\sin(\mathbf{k}\cdot\hat{\mathbf{X}}\xi\tau/2)}{\mathbf{k}\cdot\hat{\mathbf{X}}/2}\cos\left(\mathbf{k}\cdot\mathbf{x}_0+\mathbf{k}\cdot\dot{\hat{\mathbf{X}}}v\tau\right), \tag{II–2.3}$$

where v and ξ are the string network velocity and comoving correlation length, defined in section II–2.1.2 below, and $\mathbf{x}_0$ is the position of endpoint of a string segment. The string segment is parametrised by

$$\mathbf{x}(\sigma,\tau) = \mathbf{x}_0+\sigma\hat{\mathbf{X}}+v\tau\dot{\hat{\mathbf{X}}}, \tag{II–2.4}$$

with the string orientations and velocity orientations

$$\hat{\mathbf{X}} = \begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix}, \tag{II–2.5}$$

$$\dot{\hat{\mathbf{X}}} = \begin{pmatrix}\cos\theta\cos\phi\cos\psi-\sin\phi\sin\psi\\ \cos\theta\sin\phi\cos\psi+\cos\phi\sin\psi\\ -\sin\theta\cos\psi\end{pmatrix}. \tag{II–2.6}$$

$\dot{\hat{\mathbf{X}}}$ is transverse to $\hat{\mathbf{X}}$ such that $\hat{\mathbf{X}}\cdot\dot{\hat{\mathbf{X}}} = 0$. Note that the position of the string endpoint appears only through a phase in the cosine factor in equation (II–2.3), denoted $\chi \equiv \mathbf{k}\cdot\mathbf{x}_0$. The other components of the string energy-momentum tensor are given by

$$\Theta_{ij} = \left(v^2\dot{\hat{\mathbf{X}}}_i\dot{\hat{\mathbf{X}}}_j-\frac{1-v^2}{\alpha^2}\hat{\mathbf{X}}_i\hat{\mathbf{X}}_j\right)\Theta_{00}, \tag{II–2.7}$$

with $i, j = 1, 2, 3$. Choosing coordinates so that $\mathbf{k}$ lies along the $\hat{k}_3$ axis, the scalar, vector and tensor anisotropic stresses are given by

$$\Theta^{\mathrm{S}} = \frac{1}{2}(2\Theta_{33} - \Theta_{11} - \Theta_{22}), \tag{II–2.8}$$

$$\Theta^{\mathrm{V}} = \Theta_{13}, \tag{II–2.9}$$

$$\Theta^{\mathrm{T}} = \Theta_{12}. \tag{II–2.10}$$

2.1.2 Velocity dependent one-scale model

The velocity dependent one-scale model (VOS) equations dictate the values of the string network correlation length L, and the average velocity v, of string segments in the network [214]. There are many different approaches to addressing the evolution of the string network which can take into account different physical effects [214; 353; 363]. Here only VOS is considered. The correlation length L is the average length of string segments which, for scaling networks (that have a random walk structure), is also equal to the average string separation. The network velocity v, is the root-mean-square (RMS) velocity of these correlation-length-sized string segments averaged over all (shorter) length scales. The macroscopic evolution equations for these network parameters can be derived from the Nambu-Goto action by applying a statistical averaging procedure over the string worldsheet [248; 250; 251]. Expressed in terms of the conformal time τ they read

$$L' = (1 + v^2)\mathcal{H}L + \frac{c_{\mathrm{r}} v a}{2}, \tag{II–2.11}$$

$$v' = (1 - v^2)\left(\frac{\tilde{k}a}{L} - 2\mathcal{H}v\right), \tag{II–2.12}$$

where $' \equiv d/d\tau$, unlike in equation (II–2.2). The loop chopping efficiency parameter c_{r}, quantifies the energy loss due to loop production and $\tilde{k}$ provides a phenomenological description of the small-scale structure on the string, which, for relativistic strings, is given by

$$\tilde{k} = \frac{2\sqrt{2}}{\pi}\left(\frac{1 - 8v^6}{1 + 8v^6}\right). \tag{II–2.13}$$

Recalling that the correlation length can be written in comoving units as $\xi\tau = L/a$. The VOS equations in comoving units are

$$\xi' = \frac{1}{\tau}\left(\mathcal{H}v^2\xi\tau - \xi + \frac{c_{\mathrm{r}} v}{2}\right), \tag{II–2.14}$$

$$v' = (1-v^2)\left(\frac{\tilde{k}}{\xi\tau} - 2\mathcal{H}v\right), \tag{II–2.15}$$

For fixed expansion rate the scaling solutions, found by the requirement $\xi' = 0$ and $v' = 0$, read

$$\xi = \sqrt{\frac{\tilde{k}(\tilde{k}+c_{\rm r})(1-\beta)}{4\beta}}, \tag{II–2.16}$$

$$v = \sqrt{\frac{\tilde{k}(1-\beta)}{\beta(\tilde{k}+c_{\rm r})}}, \tag{II–2.17}$$

where β is the physical time FLRW expansion exponent $a \propto \tau^\beta$ and is equal to 1 and 2 in the radiation and matter eras respectively. Note in the scaling solutions of (II–2.17) the implicit velocity dependence of $\tilde{k}$ through equation (II–2.13). Earlier implementations of the cosmic defect CMB code `CMBACT` [288] used two sets of values for the loop chopping efficiency and the parameter $\tilde{k}$ in the scaling solutions (II–2.17) for the radiation and matter eras. These values were then interpolated between for the transition between the radiation and matter eras. However, in the latest implementation of the VOS equations in `CMBACT4` [287], the velocity dependence of $\tilde{k}$ is explicitly used and the loop chopping efficiency is kept constant throughout both epochs [248]. This approach is adopted here: at any particular τ, the values of ξ and v, found using the VOS equations (II–2.14 – II–2.15), are used for calculating the UETC, keeping $c_{\rm r}$ constant throughout and explicitly accounting for the velocity dependence (II–2.13) of $\tilde{k}$. In earlier versions of `CMBACT` the wiggliness α, was also an evolving parameter, but it is now kept constant in `CMBACT4`, which is the approach taken here. The evolution of the network parameters can be seen for a range of $c_{\rm r}$ in figure II–2.1 showing that a wide range of correlation lengths and velocities are available. Detailed comparison of the VOS model with Nambu-Goto simulations of ordinary string networks (i.e. single string type with unit intercommuting probability [330]) determine the loop chopping efficiency to $c_{\rm r} = 0.23 \pm 0.04$ [248], corresponding to the black dot-dot-dashed curves in figure II–2.1. Models of cosmic superstrings generally have suppressed intercommutation probabilities [157; 192; 193; 198], which effectively reduces $c_{\rm r}$ and so they correspond to the purple region in the figure. Such networks have relativistic RMS velocities $v \sim 1/\sqrt{2}$ and correlation lengths much smaller than the horizon, corresponding to a much higher string number density compared to ordinary string networks. However, they also have smaller string tension so their overall effect on the CMB can be small, consistent with the data.

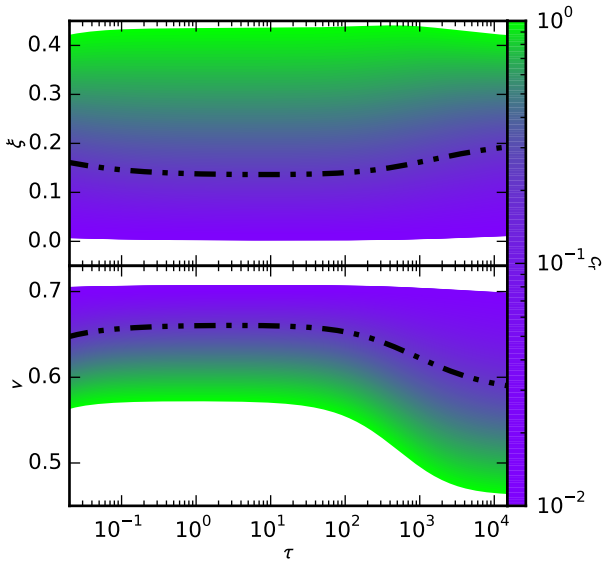


Figure II–2.1: The evolution of the velocity v, and correlation length ξ, for a range of $c_{\rm r} = [10^{-2}, 1.0]$. The black dot-dot-dash line indicates the correlation lengths and velocities obtained when $c_{\rm r} = 0.23$. The greener area (lighter in black and white) of the plot indicates larger values of $c_{\rm r}$ whilst the more purple region (darker in black and white) shows smaller $c_{\rm r}$.

It should be noted that the RMS network velocity used in the VOS model arises from a worldsheet average and is thus integrated over all (short) length scales. Therefore, it provides an accurate measure of the energy stored in a wiggly string segment, but does not explicitly correspond to (and in fact is expected to be larger than) the coherent velocity on correlation-length scales. Indeed, numerical simulations of Nambu-Goto strings reveal a network velocity distribution with larger velocities at short scales, implying that the RMS velocity is dominated by relativistic speeds at short distances. On length scales of order the correlation length, coherent velocities as low as $v_{\rm coh} \simeq 0.2$ have been reported [25; 54; 249; 252]. Other network velocity measures (again containing information from a range of length scales) in both Nambu-Goto and Abelian-Higgs string simulations also tend to be lower than the VOS RMS velocity, with velocities in the Abelian-Higgs model $v_{AH} \simeq 0.5$, significanlty slower than in Nambu-Goto simulations [60; 63; 174]. For further discussion about the impact of string velocities on the UETC and the string power spectrum see the end of Part II–2.1.6.

2.1.3 Unconnected segment model

Simulations of evolving string networks are numerically very expensive. Strings decay as $1/(\xi\tau)^3$, eventually reaching a scaling solution (ξ = constant) with a number density of tens to hundreds of strings per horizon volume. At early times, the box contains a huge number of strings whose dynamics and interactions have to be tracked at each time step. The USM [22; 288] dramatically reduces the required computational resources by approximating the string network as a collection of correlation-length-sized segments, with the time evolution of the correlation length and segment velocity described by the VOS equations. Moreover, the model consolidates these string segments by collecting all strings that decay between any two times, and so fewer strings need to be tracked. The number of strings that decay between any two conformal times in a volume V, is

$$N_{\rm d}(\tau_i) \quad = \quad V[n(\tau_{i-1}) - n(\tau_i)]\,, \tag{II–2.18}$$

where $n(\tau)$ is the number density of strings at conformal time τ, given by $n(\tau) = C(\tau)/(\xi\tau)^3$. In `CMBACT`, the factor $C(\tau)$ is chosen so as to keep the number of strings at any time proportional to $1/(\xi\tau)^3$. The energy-momentum tensor for the string network is then given by the sum over the total number of consolidated string segments K , with a factor accounting for string decay

$$\Theta_{\mu\nu} \quad = \quad \sum_{i=1}^{K} \sqrt{N_{\rm d}(\tau_i)}\Theta^i_{\mu\nu} T^{\rm off}(\tau, \tau_i, L_{\rm f}). \tag{II–2.19}$$

The string decay factor $T^{\rm off}(\tau, \tau_i, L_{\rm f})$ is a function interpolating between 1 and 0 and is responsible for turning off the contribution of the $i^{\rm th}$ consolidated segment after the time it has decayed. Its steepness is controlled by a string decay parameter $0 < L_{\rm f} \leq 1$, as follows:

$$T^{\rm off}(\tau, \tau_i, L_{\rm f}) \quad = \quad \begin{cases} 1 & \tau < L_{\rm f}\tau_i \\ 1/2 + 1/4(y^3 - 3y) & L_{\rm f}\tau_i < \tau < \tau_i \\ 0 & \tau_i < \tau \end{cases} \tag{II–2.20}$$

where

$$y \quad = \quad \frac{2\ln(L_{\rm f}\tau_i/\tau)}{\ln(L_{\rm f})} - 1\,. \tag{II–2.21}$$

Thus, in the limit $L_{\rm f} \to 1$ the string decay factor $T^{\rm off}(\tau, \tau_i, L_{\rm f})$ approaches a Heaviside function, sharply switching off the contribution of the $i^{\rm th}$ consolidated segment to the network energy-momentum tensor for times $\tau > \tau_i$.

The $L_{\rm f}$ parameter

Since the number of consolidated segments also sets the number of decay epochs, a finite number of consolidated segments leads to discrete steps in the number density of strings. The string decay parameter $L_{\rm f}$ was introduced to allow a fraction of the consolidated strings to decay before the end of their respective decay epoch, thus making the number density evolution smoother. The function $C(\tau)$ was also introduced to ensure that the number of strings at any conformal time τ is kept proportional to $(\xi\tau)^{-3}$. However, one consequence of $L_{\rm f} < 1$ is that it is possible that $L_{\rm f}\tau_{i+1} < \tau_i$, meaning strings can start to decay earlier than their respective epoch and the number density is systematically lower.

In the `CMBACT4` implementation changing the number of consolidated segments from 200 to 10000 has very little impact on the string spectra, as shown in figure II–2.2. However, the amplitude of the C_ℓ *is* dependent on the value of $L_{\rm f}$. The change is scale dependent, but can be as much as 30%, for example near the peak of the scalar temperature signal. Previous analyses which have used the results from `CMBACT` have overlooked this dependence. Although not entirely degenerate with the amplitude of C_ℓ, which scales proportional to $(G\mu)^2$, it will clearly have some affect on the inferred values of $G\mu$ from the USM. This approach is compared in the following section.

Infinite consolidated string segments

A large number of segments can be accommodated analytically. As discussed in [37], the scaling factor, that weights the UETC taking into account string decay, has a particularly simple form when the number of consolidated string segments tends to infinity, $L_{\rm f} \to 1$ and $C(\tau) \to 1$. This is

$$\begin{aligned} f(\tau_1,\tau_2,\xi(\tau_1)\,,\xi(\tau_2)) &= \sum_{i=1}^{K} N_{\rm d}(\tau_i) T^{\rm off}(\tau_1,\tau_i,L_{\rm f}) T^{\rm off}(\tau_2,\tau_i,L_{\rm f}), \\ &= (\xi(\mathrm{Max}[\tau_1,\tau_2])\mathrm{Max}[\tau_1,\tau_2])^{-3}\,, \\ &= f\big(\tau_{\rm Max},\xi(\tau_{\rm Max})\,.\big) \end{aligned} \qquad \text{(II–2.22)}$$

An analytic expression for the scaling factor can also be found for arbitrary $L_{\rm f}$ using the form of $T_{\rm off}$ quoted in equation (II–2.20). However, it seems natural to consider only the case $L_{\rm f} = 1$ when the number of consolidated string segments is very large. In the infinite limit the segments will decay at an infinite number of epochs which are infinitesimally separated, a continuous limit in which the string number density is smooth. The number density scales according to $(\xi\tau)^{-3}$ in this approach. While infinite consolidated segments may seem unphysical, it is just a

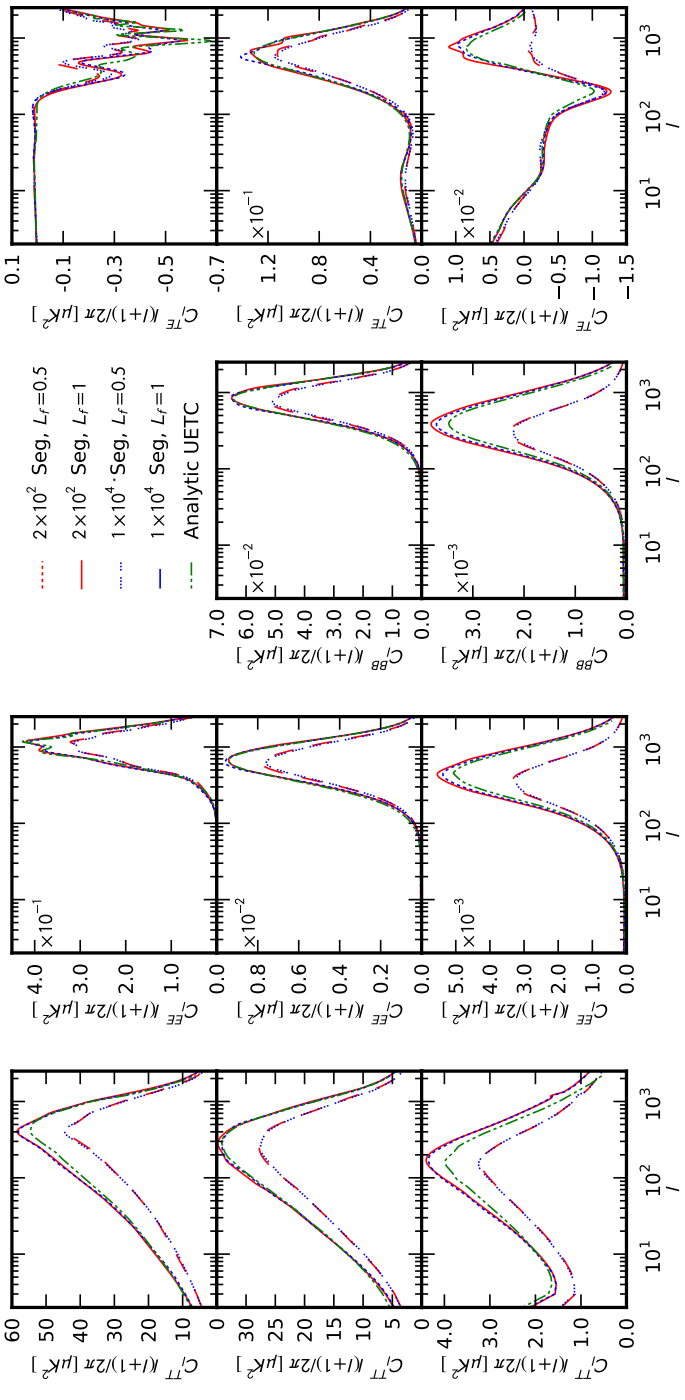


Figure II–2.2: C_ℓ obtained from the string realisation code CMBACT4 with 200 and 10000 consolidated string segments for 2000 string realisations between the red solid and dashed lines and blue dotted and dotted-segment lines respectively. The solid red and dotted blue lines at the top of each band indicate a value of $L_{\rm f} = 1$ for 200 and 10000 segments, while the red dotted and blue dotted-segment lines show $L_{\rm f} = 0.5$. The top, middle and bottom rows show the scalar, vector and tensor C_ℓ modes respectively. The first column contains the temperature (TT) C_ℓ, the second column has the EE mode contribution, BB modes are in the third and the TE cross-correlation in the final column. We also plot the corresponding spectra derived from our analytic USM method, shown in green dot-dot-dashed lines.

limit used to obtain the correct scaling relation. Very similar results to `CMBACT4` are obtained when using between 200 to 10000 segments with $L_{\rm f} = 1$. The question of whether the observed resulting modification of scaling from early string decay obtained when $L_{\rm f} < 1$ is physical or not requires investigation. Since $C(\tau) = 1$, different scaling behaviour does not need to be considered. Ultimately, the USM is a simplified model which aims to match the UETC from simulations by adjusting the network parameters. Overall it has been shown to match Nambu-Goto simulations well [227]. However, due to the correlation between the inferred values for $G\mu$ for a given $L_{\rm f}$, this issue should be considered more closely. Since the number density scales according to $(\xi\tau)^{-3}$ using the approach here, it is adopted for the comparison to data.

2.1.4 Analytic calculation of the unequal-time correlator

The UETC can be computed analytically [37] by integrating over all string configurations (orientations and positions) in the network. For the two point correlator between $\Theta(\tau_1, \mathbf{k}_1, \chi_1)$ and $\Theta(\tau_2, \mathbf{k}_2, \chi_2)$ translational invariance implies $\mathbf{k}_1 = -\mathbf{k}_2 = \mathbf{k}$ and so $\chi_1 = -\chi_2 = \chi$. Considering that, due to equations (II–2.3) and (II–2.7), $\Theta(\tau, \mathbf{k}, \chi)$ is a symmetric function of $\mathbf{k}$ the integral is

$$\begin{aligned}\langle\Theta(\tau_1,\mathbf{k})\Theta(\tau_2,\mathbf{k})\rangle &= \frac{2f(\tau_{\rm Max},\xi(\tau_{\rm Max}))}{16\pi^3}\int_0^{2\pi}d\phi\int_0^{2\pi}d\psi\int_0^{\pi}\sin\theta d\theta \\ &\quad\times\int_0^{2\pi}d\chi\Theta(\tau_1,\mathbf{k},\chi)\Theta(\tau_2,\mathbf{k},\chi). \qquad \text{(II–2.23)}\end{aligned}$$

Without loss of generality $\mathbf{k}$ can be chosen to lie along the k_3-axis, such that $\mathbf{k} = k\hat{k}_3$. Θ here represents each of Θ_{00}, $\Theta^{\rm S}$, $\Theta^{\rm V}$ and $\Theta^{\rm T}$ in equations (II–2.8–II–2.10). The ϕ, ψ and χ integrals can be done analytically in this case leaving only the θ integral in terms of Bessel functions. The UETC can then be written as the sum over six integral identities

$$\begin{aligned}\langle\Theta(\tau_1,k)\Theta(\tau_2,k)\rangle &= \frac{f(\tau_{\rm Max},\xi(\tau_{\rm Max}))\mu^2}{k^2\sqrt{1-v(\tau_1)^2}\sqrt{1-v(\tau_2)^2}} \\ &\quad\times\sum_{i=1}^{6}A_i[I_i(x_-,\varrho)-I_i(x_+,\varrho)], \qquad \text{(II–2.24)}\end{aligned}$$

where $\varrho = k|v(\tau_1)\tau_1 - v(\tau_2)\tau_2|$ and $x_\pm = (x_1 \pm x_2)/2$ with $x_{1,2} = k\xi(\tau_{1,2})\tau_{1,2}$. Here $x_{1,2}$ means x_1 or x_2 respectively. This extends the corresponding result of [37] in that ξ and v are now functions of τ instead of being kept constant. This means that the expressions of the amplitudes A_i, presented in Table II–A.1 in Appendix II–A, are now time-dependent. The integral identities (shown in equations (II–A.1)–(II–A.6) in Appendix II–A) remain the same. It should be noted that $I_1(x,\varrho)$ and $I_4(x,\varrho)$ diverge but the combination $I_{1,4}(x_-,\varrho) - I_{1,4}(x_+,\varrho)$ is

regular and, in the limit where $x_{1,2} \gg x_{2,1}$, has an analytic approximation given by

$$I_1(x_-, \varrho) - I_1(x_+, \varrho) = \frac{\pi x_{1,2}}{2} J_0(\varrho), \tag{II–2.25}$$

$$I_4(x_-, \varrho) - I_4(x_+, \varrho) = \frac{\pi x_{1,2}}{2\varrho} J_1(\varrho). \tag{II–2.26}$$

In the small $x_{1,2}$, limit the UETC can be written as

$$\langle \Theta(\tau_1, k)\Theta(\tau_2, k)\rangle = \frac{f(\tau_{\mathrm{Max}}, \xi(\tau_{\mathrm{Max}}))\mu^2}{k^2\sqrt{1-v(\tau_1)^2}\sqrt{1-v(\tau_2)^2}} B, \tag{II–2.27}$$

and at equal times, when $x_1 = x_2 = x$ and $\varrho = 0$, the equal-time correlator is given by

$$\langle \Theta(\tau, k)\Theta(\tau, k)\rangle = \frac{f(\tau, \xi(\tau))\mu^2}{k^2(1-v(\tau)^2)} C. \tag{II–2.28}$$

The form of B and C are similar to [37] but again depend on the values of v and ξ at τ_1 and τ_2. These coefficients have been included in Table II–A.2 in Appendix II–A. Thanks to these analytic approximations, computational times can be greatly reduced compared to the case where the integral identities I_i are used for computation over the whole range of $k\tau_1$, $k\tau_2$. The regions where these approximations are valid are shown in figure II–2.3, only the white region is computationally intensive. It should be noted that, because ξ is a function of time, the shape of the approximated regions in figure II–2.3 changes for different values of k and so a large number of k-modes must be considered when computing the UETC. This is in contrast to [37], where the approximation of constant ξ and v meant that the UETC was only a function of the combinations $k\tau_1$ and $k\tau_2$.

Negative values of the UETC

It has been noted in [297] that there are negative regions in the string UETC calculated analytically through the formalism used here, which do not appear in the Gaussian model for the string UETC used in [297]. These can be seen in figure II–2.4.

There are two distinct types of regions with negative values of the UETC. First, regions with small $k\tau_1$ and large $k\tau_2$ (and vice versa), corresponding to the top left and bottom right corners of figure II–2.3 or figure II–2.4: in these regions the UETC should be zero, but small negative (and positive) values can arise from the finite order truncation of the Bessel series expansions of $I_1(x_\pm, \rho)$ and $I_4(x_\pm, \varrho)$ in equation (II–2.24). These values are spurious and can be thought

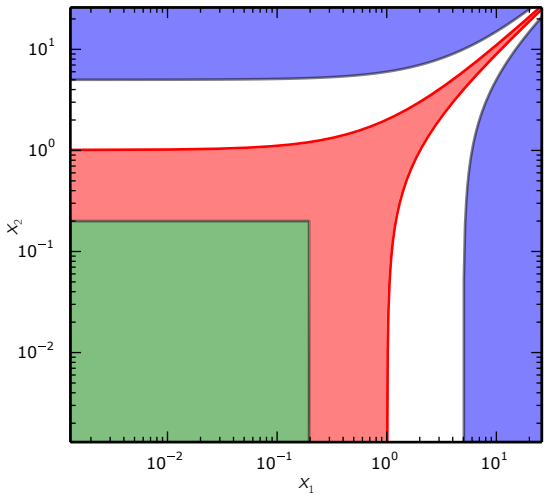


Figure II–2.3: The regions of $x = k\tau\xi$ covered by analytic approximations. In green is the region when $x_1 \ll 1$ and $x_2 \ll 1$, red when $|\log x_1 - \log x_2| < \epsilon$ and blue when $|x_1 - x_2| \gg 1$. In the code the $x_{1,2} \ll 1$ region is set for $x_{1,2} < 0.2$, $\epsilon = 0.001$ for $x_1 \approx x_2$ and $|x_1 - x_2| > 10$ for $x_{1,2} \gg x_{2,1}$.

of as noise arising from the truncation. The order of truncation must then be chosen such that this noise is at a tolerable level.

Second, in the regions off the diagonal with large $k\tau_1 \approx k\tau_2$ (corresponding to the top right corner of figure II–2.3 or figure II–2.4) there is a ringing pattern with successive positive and negative peaks that decay away from the diagonal. These oscillatory patterns are a consequence of causality [22; 116; 359], built into the USM: as the correlator must vanish at superhorizon scales (in fact in the USM it vanishes at scales larger than the correlation length, which is smaller than the horizon), this introduces a sharp edge in physical space that becomes oscillatory in Fourier space. This oscillatory pattern therefore has a clear physical origin, but in the USM it is somewhat artificially enhanced due to the fact that the model assumes all string segments have the same length. If segments are instead given a length distribution peaking at the network correlation length, the sharp edge is smoothed and the oscillatory pattern gets suppressed. Further, considering a segment velocity distribution peaking near the network RMS velocity again suppresses these oscillations. The Gaussian model assumes a wide Gaussian distribution of string lengths (but also assigns non-zero values to the correlator at superhorizon scales) so this causal oscillatory feature is absent from the UETC in that model.

The suppression of oscillations in the UETC can be seen in figure II–2.5 where

the blue solid line shows the profile of the UETC across the diagonal as calculated using the velocity and correlation lengths from VOS. In red dot-dot-dash is the same profile when a Gaussian distributed sample of velocities and correlation lengths, peaking on the VOS values, are chosen. The oscillatory features are mostly washed out but the first trough remains a prominent feature. The off-diagonal dip in the correlation functions that are found after considering a range of segment lengths and velocities has also been observed in Abelian-Higgs simulations [60]. It may also be related to the velocity anti-correlation observed in Nambu-Goto simulations on correlation-length scales and can be attributed to string intercommutations [249].

2.1.5 Eigenmode decomposition

The UETC is generally rescaled by a factor of $\sqrt{\tau_1\tau_2}$, which, for ξ and v constant, makes it a function of $k\tau_1$ and $k\tau_2$ only. This is not true in the present case because now the time-dependence of ξ and v is tracked, so the UETC depends separately on k, τ_1 and τ_2. However, it is still useful to introduce this rescaling in order to facilitate direct comparison of the UETC with previous results. This rescaled UETC can then be discretised onto a logarithmic grid in $k\tau_1$ and $k\tau_2$ with $n \times n$ grid points and then diagonalised giving the eigenvectors and eigenvalues ([283])

$$(k^2\tau_1\tau_2)^\gamma\sqrt{\tau_1\tau_2}\langle\Theta(\tau_1,k)\Theta(\tau_2,k)\rangle \quad = \quad \sum_{i=1}^{N}\lambda_i u_i(k\tau_1)\otimes u_i(k\tau_2). \quad \text{(II–2.29)}$$

Due to the explicit dependence on k, this diagonalisation procedure has to be repeated for a large number of k-modes, and the eigenvalues are k-dependent. This significantly increases the computation time compared to [37]. The extra factor $(k^2\tau_1\tau_2)^\gamma$ is used for more efficient reconstruction of the UETC when the eigenmodes are truncated below n. The choice $\gamma = 0.25$ gives the best reconstruction on scales that give the dominant contribution to the CMB anisotropies.

There is no correlation between the scalar, vector and tensor modes and so the vector and tensor UETC can be diagonalised independently. However, the density Θ_{00}, and scalar anisotropic stress $\Theta^{\rm S}$, are correlated. The diagonalisation is done over a $2n \times 2n$ grid constructed from

$$\begin{array}{c|c} \langle\Theta_{00}(\tau_1,k)\Theta_{00}(\tau_2,k)\rangle & \langle\Theta^{\rm S}_{00}(\tau_1,k)\Theta^{\rm S}_{00}(\tau_2,k)\rangle \\ \hline \langle\Theta^{\rm S}_{00}(\tau_1,k)\Theta^{\rm S}_{00}(\tau_2,k)\rangle & \langle\Theta^{\rm S}(\tau_1,k)\Theta^{\rm S}(\tau_2,k)\rangle \end{array}, \qquad \text{(II–2.30)}$$

where $\langle\Theta^{\rm S}_{00}(\tau_1,k)\Theta^{\rm S}_{00}(\tau_2,k)\rangle$ is the symmetric combination of the cross-correlation between Θ_{00} and $\Theta^{\rm S}$. After diagonalisation, the first half of the eigenvectors

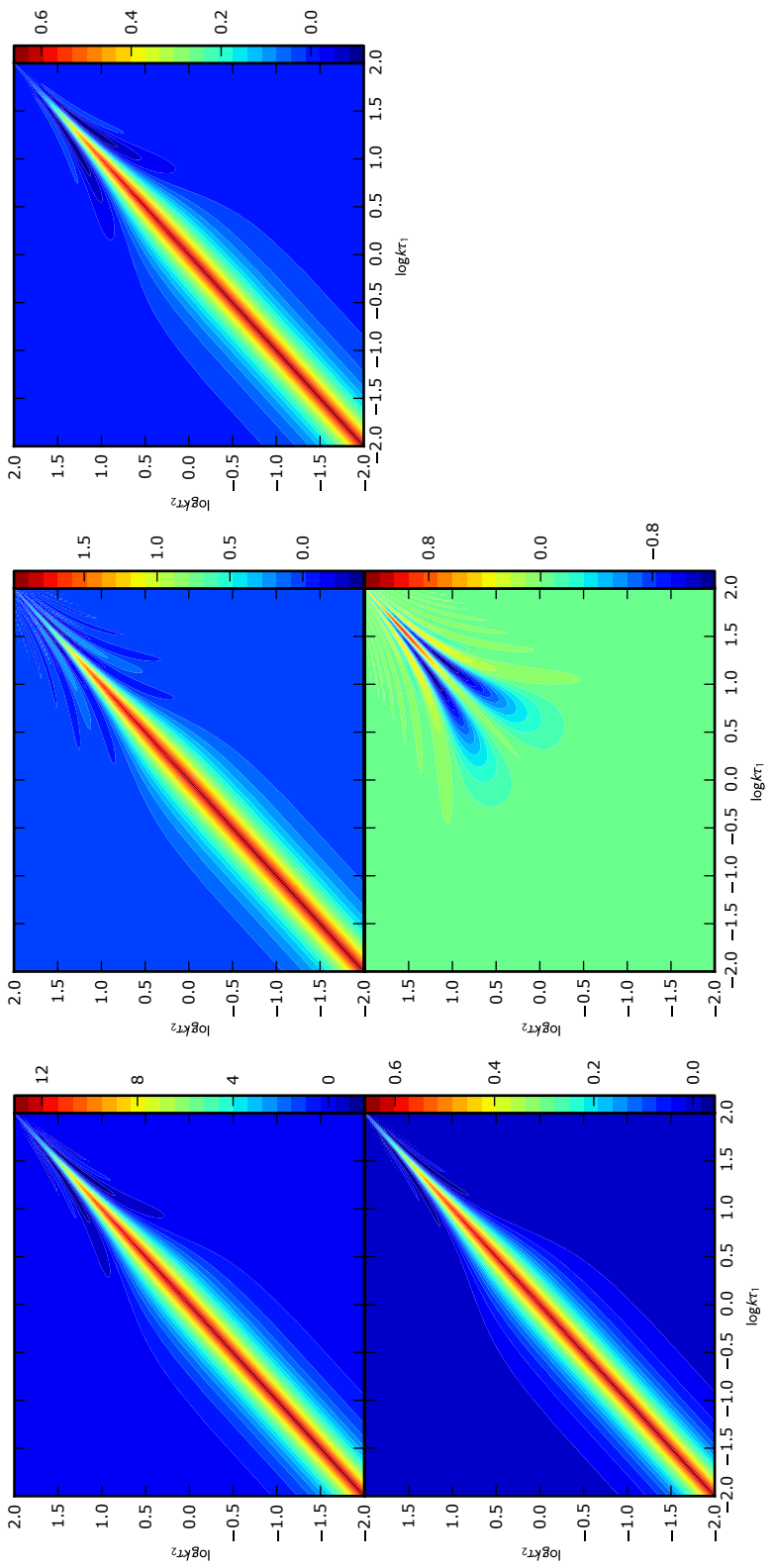


Figure II-2.4: The UETC calculated at $k = 0.05h$/Mpc. On the top row, from left to right, the subplots show 00-component followed by the scalar and vector anisotropic stresses. The bottom subplots, left to right, are the tensor anisotropic stress and the cross-correlation between the energy-density and the scalar anisotropic stress.

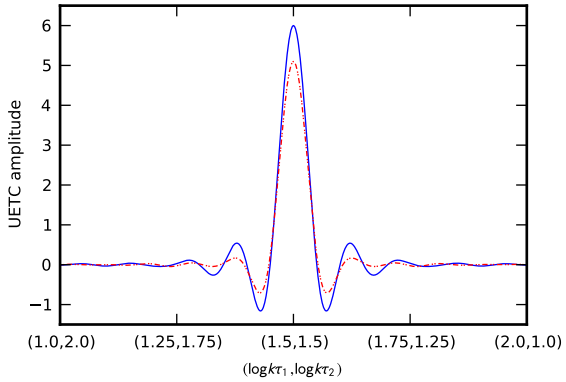


Figure II–2.5: Profile of the UETC across the diagonal in the oscillatory region with large $k\tau_1 \approx k\tau_2$. The solid blue line shows the amplitude of the UETC using the value of the velocity and correlation length from the VOS equations whilst the red dot-dot-dash lines has Gaussian distributed velocities and correlation lengths about the VOS values.

refer to the density and the second to the anisotropic stress. The diagonalisation creates orthogonal eigenvectors which are then used as source terms in the `CAMB` [237] linear Einstein-Boltzmann code. The C_ℓ are calculated using each individual eigenvector $u_i(k\tau)/(\sqrt{\tau}(k\tau)^\gamma)$, as a source function C_ℓ^i, which can be summed to give the total power spectra

$$C_\ell \quad = \quad \sum_{i=1}^{n} \lambda_i C_\ell^i. \tag{II–2.31}$$

By ordering λ_i from largest to smallest, the required accuracy in the C_ℓ can be achieved by including relatively few eigenmodes. This can be seen in the middle row of figure II–2.6 where there is only about 10% difference between using all 512 eigenmodes of a 512×512 grid compared to only using 32 eigenmodes when fixing the value of $G\mu$. Also, it can be seen in the top row of figure II–2.6 that reducing the grid resolution reduces the amplitude of the C_ℓ. A grid resolution of 128×128 is about 5% lower, on average, than using the 512×512 grid but convergence times decrease drastically. It should be noted that there is negligible difference between using a 512×512 and a 1024×1024 grid meaning that the former is reliably giving the full C_ℓ contribution. The bottom row shows what happens when using more k values in the calculation. Wiggly features arise from using too few k values and can be removed at the expense of a much longer calculation. These findings are used to choose the optimal UETC parameters to give good quality C_ℓ in a reasonable amount of time. The resulting spectra obtained from our analytical method are shown in figure II–2.2 in green dot-dot-dashed curves and agree well with USM string realisations, especially in the

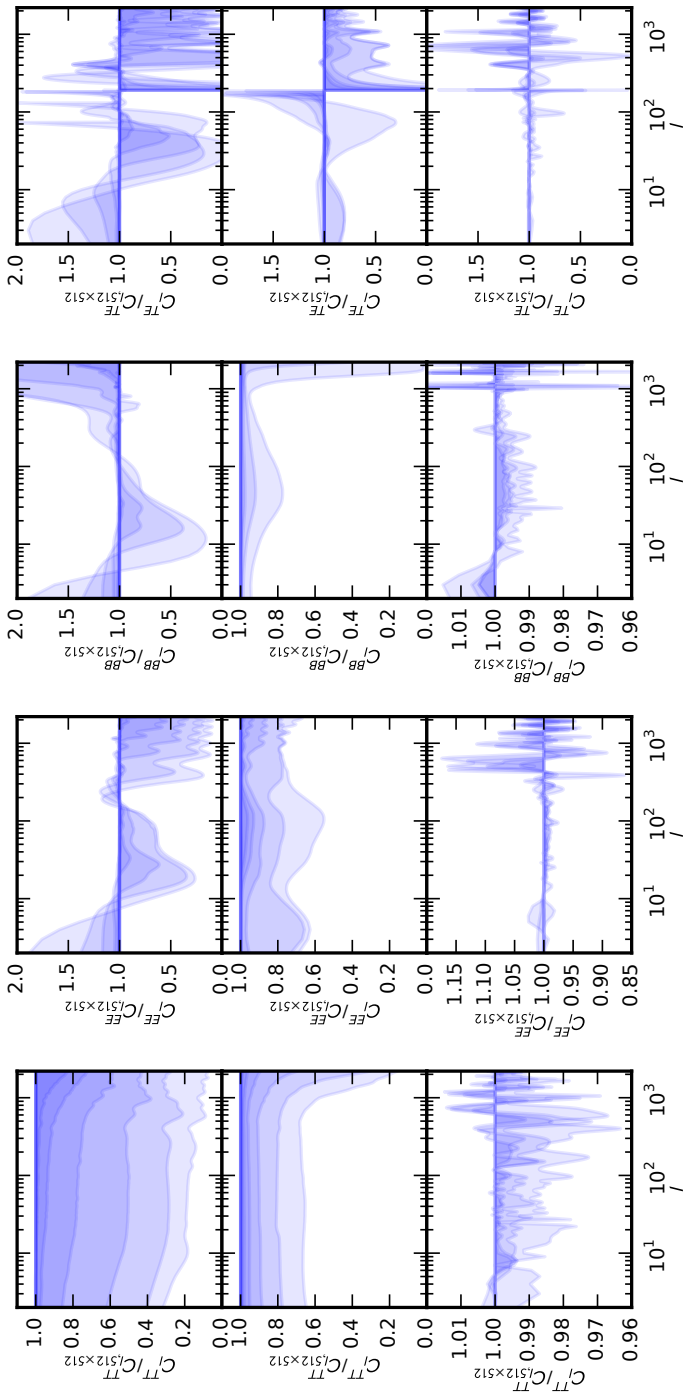


Figure II–2.6: The ratio of the C_ℓ calculated using a UETC with a 512×512 grid with all the eigenmodes and: (Top row) other grid resolutions, the lightest shaded region with an 8×8 grid whilst the darkest a 256×256 grid; (Middle row) when fewer eigenmodes are included, 8 eigenmodes for the lightest shaded region and 256 for the darkest; (Bottom row) when the `accuracy_boost` setting of `CAMB` is decreased, reducing the number of k values.

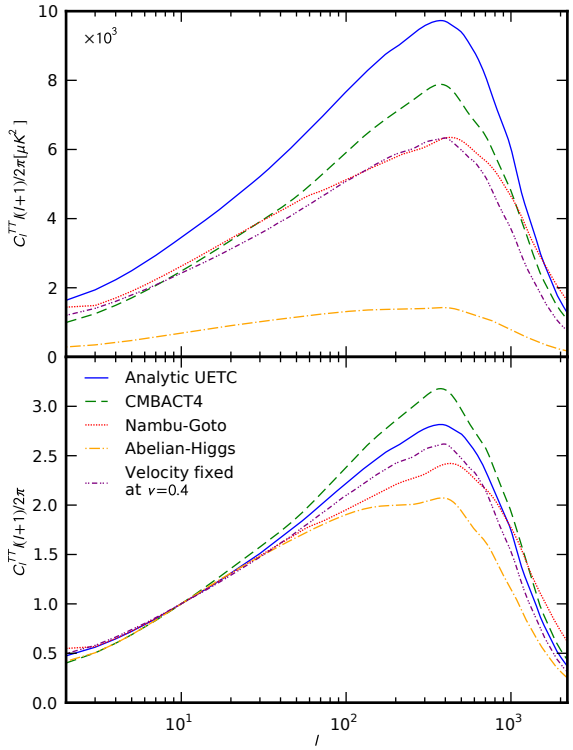


Figure II–2.7: Comparison of approaches to string modelling, scaled by $G\mu$ in the upper subplot and normalising the temperature power spectrum at $\ell = 10$ in the lower subplot. The approach developed here (in solid blue) is compared to `CMBACT4` [288], Nambu-Goto simulations [227], and Abelian-Higgs simulations [60] (in dashed green, dotted red, dot-dashed orange and the analytic USM with the velocity fixed at $v = 0.4$ in dot-dot-dashed purple respectively).

limit of large numbers of simulated segments.

2.1.6 Comparison of the string power spectrum

In figure II–2.7 the temperature power spectrum calculated here (scaled by $G\mu$ in the upper subplot and normalised at $\ell = 10$) is compared to those of `CMBACT4` [288], Nambu-Goto simulations [227], and Abelian-Higgs simulations [60]. Both `CMBACT4` and this method use the same velocity dependent one-scale model parameters, but `CMBACT4` uses $L_{\rm f} = 0.5$. The Nambu-Goto simulations are performed in an expanding background from recombination to today, including Λ domination. Large loops are kept in the simulation and contribute to the total energy-momentum tensor of the network, but these simulations cannot resolve small-scale physics near the string width and do not include the effects of radiation backreaction. In contrast, the Abelian-Higgs simulations can resolve small-scale structure and radiative effects [101]. These, however, have smaller dynamical range and cannot easily evolve through the radiation-matter transi-

tion (so the UETC is instead interpolated), but see recent progress in [101] where the authors simulate through the transition.

In summary, given the large differences in modelling between the various approaches means this comparison is encouraging, although more work is needed to further delineate the differences. In particular, as discussed at the end of Part II–2.1.2, the VOS RMS velocity is defined through a worldsheet integral over all scales and receives a large contribution from relativistic wiggles on the string. On the other hand, the USM assumes straight segments moving at a given speed and the small-scale structure on the segments is captured via a "renormalisation" of their tension. This implies that the speed to be associated to the USM segments must be lower than the VOS RMS velocity, and should correspond to the network velocity at correlation length scales. Numerical simulations show this to be significantly lower than the RMS speed. This issue has not been examined before, partly because the calculated string spectra from different approaches can differ by up to a factor of two, and partly because it can be offset by choosing a lower value for the USM parameter $L_{\rm f}$ (see below). As quantitative agreement between the different approaches is now being established, it is important to fully understand this issue. To this end it will be important to extract the network velocity distribution as a function of length scale in both Nambu-Goto and Abelian-Higgs simulations.

Plotted in figure II–2.7 in purple dot-dot-dash is the C_ℓ obtained when $v = 0.4$. As can be seen, the peak of the velocity fixed C_ℓ has a very similar amplitude to the Nambu-Goto simulation C_ℓ in dotted red, although the simulations still have larger power at both lower and higher ℓ. This supports the idea that the discrepancy in the amplitude of string spectra could be related to different predictions/assumptions on string velocity in the different approaches (cf. discussion at the end of Part II–2.1.2). Note that the parameter $L_{\rm f}$ in the USM is somewhat degenerate with the string velocity - for fixed v a lower $L_{\rm f}$ reduces the density of strings by increasing the string decay rate, thus reducing the C_ℓ amplitude and matching simulations better than using $L_{\rm f} = 1$. In the absence of a more complete quantitative understanding of the string velocity distribution - input required from string evolution simulations - the string spectra obtained from the USM have a larger amplitude (see the solid blue line in the upper subplot of figure II–2.7). This leads to slightly tighter constraints on cosmic strings than in numerical simulations. Marginalising over the network parameters $c_{\rm r}$ and α, partly takes care of the differences between $L_{\rm f} = 0.5$ and $L_{\rm f} = 1$ in the USM since high $c_{\rm r}$ reduces the velocity (as seen from equation (II–2.15) and pictorially

in figure II–2.1).

Overall, when normalised at $\ell = 10$, the four spectra agree reasonably well. The USM variants (`CMBACT4` and the approach here) both predict slightly more power at the peak than either of the simulations. The Nambu-Goto simulations predict more power on very small scales, around twice as much as the Abelian-Higgs model. It is well known that Nambu-Goto calculations yield higher string densities than field theoretic ones, which will increase their overall normalisation. The resulting constraints on $G\mu$ are therefore around a factor of 50% lower [14] as can be inferred from the upper subplot in figure II–2.7. The USM variants are closer to the Nambu-Goto simulations in this respect [227]. Within this work, using the analytic USM to mimic the Abelian-Higgs spectra is not considered. As shown, there is some additional uncertainty in the USM, as the normalisation depends somewhat on the choice of L_{f}.

2.2 Cosmic superstrings

A cosmic superstring network can be modelled as a collection of string segments of different types, each string type having its own tension and self-intercommuting probability [34; 35; 92; 157; 173; 192; 193; 198; 293; 360; 361]. Strings of different types interact with each other via "zipping" or "unzipping" leading to heavier or lighter strings respectively, that are connected to the original strings at trilinear Y-shaped junctions [291]. The fundamental building blocks for these networks are light (fundamental) F-strings and heavier (Dirichlet) D-strings, with a tension hierarchy controlled by the fundamental string coupling [291; 322; 379]. Heavier strings arise as bound states between p F-strings and q D-strings, where p,q are coprime. Given the fundamental string tension, the corresponding tensions of these heavier (p, q)-strings are controlled mainly by p,q and the value of the string coupling. These networks generally behave very differently than their ordinary cosmic string counterparts. They are typically characterised by small intercommutation probabilities, thus leading to higher string number densities [34; 198; 293; 360]. The complex interactions present imply that several string types with different tensions and correlation lengths can simultaneously contribute to the string network CMB spectra.

In scaling superstring networks, the string number density is dominated by the lightest F-strings, followed by D-strings and the first bound state, i.e. (1,1)-strings. Heavier bound states are suppressed, so the number of string types considered in the model can be truncated at a finite number. Following [293] the

network is described by keeping seven distinct types of strings:

1	F	$(1,0)$,	
2	D	$(0,1)$,	
3	FD	$(1,1)$,	
4	FFD	$(2,1)$,	
5	FDD	$(1,2)$,	
6	$FFFD$	$(3,1)$,	
7	$FDDD$	$(1,3)$,	(II–2.32)
$\vdots$	$\vdots$	$\vdots$	

where the last column describes the (p,q) charges of the corresponding string type.

The large-scale dynamics is then modelled by seven copies of the VOS equations, appropriately extended to account for transfer of energy among the different string types through zipping and unzipping interactions [35; 360]. In each copy of the VOS equations describing a single string, say of type i, the self interaction coefficient $c_{\rm r}$ in equation (II–2.14) is replaced by the corresponding self-interaction coefficient c_i, and new cross-interaction terms with coefficients d_{ij}^k are added to describe zipping and unzipping. The coefficients c_i, d_{ij}^k are controlled by the corresponding microphysical intercommuting probabilities $\mathcal{P}_{ij}$ [293], which can be estimated [192; 193] from the corresponding string theoretic amplitudes (and field theory approximations in the case of non-perturbative interactions between heavy strings [157]). They can be expressed as a product of two pieces: one that is dependent on the volume of the compact extra dimensions $\mathcal{V}_{ij}(w,g_{\rm s})$, and a quantum interaction piece $\mathcal{F}_{ij}(v,\theta,g_{\rm s})$. Physically, $\mathcal{V}_{ij}$ can be thought of as arising from string position fluctuations around the minimum of a localising potential well, giving rise to an effective volume seen by each type of string. The heavier the string the smaller the fluctuations are and so the smaller the value of $\mathcal{V}_{ij}$ [193]. The parameter w corresponds to the effective volume in the compact extra dimensions seen by F-strings. $g_{\rm s}$ is the fundamental string coupling and v and θ are the relative velocity and angle of the incoming strings. For a pair of strings colliding at an angle θ, and relative speed v, the intercommuting probability is

$$\mathcal{P}_{ij}(v,\theta,w,g_{\rm s}) \;=\; \mathcal{F}_{ij}(v,\theta,g_{\rm s})\mathcal{V}_{ij}(w,g_{\rm s}). \qquad \text{(II–2.33)}$$

Explicit forms for $\mathcal{F}_{ij}$ and $\mathcal{V}_{ij}$ are calculated in [293]. Since the network contains a large number of individual strings with a range of velocities and orientations,

the coefficients c_i and d_{ij}^k are determined by the integral of $\mathcal{P}_{ij}$ over a Gaussian velocity distribution centred on the scaling network velocities of each string type and over all angles. This gives the average intercommuting probabilities $\mathcal{P}_{ij}(w, g_{\rm s}) \equiv P_{ij}$. Numerical simulations of single-type Nambu-Goto strings with small intercommuting probability [34] suggest that the self-interaction coefficients c_i scale as

$$c_i \quad = \quad c_{\rm s} \times P_{ii}^{1/3}, \tag{II–2.34}$$

where $c_{\rm s}$ is the standard self-interaction coefficient in three dimensions corresponding to the value $c_{\rm r}$ in Part II–2.1.2. This choice of $c_{\rm s}$ implies a convenient normalisation of the coefficients c_i so that the ordinary cosmic string value $c_{\rm r}$ is recovered when $P_{ii} = 1$. This facilitates direct comparison with ordinary cosmic strings.

For cross-interactions between two strings of types i and j $(i \neq j)$, producing a segment of type k, there is an additional factor arising from the kinematic constraints of Y-junction formation [88; 94] that is denoted as S_{ij}^k $(i \neq j)$. This also arises as an integral over relative velocities and string orientations [33; 293]

$$\begin{aligned} S_{ij}^k \quad = \quad & \frac{1}{\mathcal{S}} \int_0^1 v^2 dv \int_0^{\pi/2} \sin\theta d\theta \\ & \times \Theta(-f_{\vec{\mu}}(v,\theta)) \exp[(v - \bar{v}_{ij})^2/\sigma_v^2] \end{aligned} \tag{II–2.35}$$

where $\mathcal{S}$ is a normalisation factor [293], $\Theta(-f_{\vec{\mu}}(v,\theta))$ imposes the kinematic constraints ($f_{\vec{\mu}}(v,\theta)$) is the condition for junction formation to be kinematically possible) [94] and σ_v^2 is the variance of the velocity distribution peaked on the relative scaling velocities $\bar{v}_{ij} = (v_i^2 + v_j^2)^{1/2}$ between strings of type i and j. The cross-interaction coefficients are then given by

$$d_{ij}^k \quad = \quad d_{ij} \times S_{ij}^k \tag{II–2.36}$$

where $d_{ij} = \kappa \times P_{ij}^{1/3}$. The overall normalisation κ is the analogue of $c_{\rm s}$, but for cross-interactions. There is no obvious choice for this phenomenological parameter, but it may be expected to be of order unity by analogy to the ordinary self-interacting string result for $c_{\rm r}$, obtained by numerical simulations. Strictly speaking it should be treated as an extra parameter for the model but, given the large computational resources required in the MCMC analysis, it is set to unity in this work. The analysis will still indirectly capture the effects of changing this parameter as it is somewhat degenerate with w. To see this, note that d_{ij} is also proportional to $P_{ij}^{1/3}$ which depends weakly on w through the volume factor $\mathcal{V}_{ij}(w, g_{\rm s})$. The leading effect of w is to change the relative amplitude

between self-interactions (FF interactions having the strongest w dependence) and cross-interactions of heavy strings, thus mimicking somewhat the effect of varying κ relative to $c_{\rm s}$. As computational power improves and this methodology is refined, κ should be re-introduced as an additional MCMC parameter.

The modified VOS equations [35; 293], in comoving units, are

$$\xi_i' = \frac{1}{2\tau}\Bigg[2v_i^2\xi_i\tau\mathcal{H} - 2\xi_i + c_i v_i + \sum_{a,b}\left(\frac{d^b_{ia}\bar{v}_{ia}\xi_i\ell^b_{ia}}{\xi_a^2} - \frac{d^i_{ab}\bar{v}_{ab}\xi_i^3\ell^i_{ab}}{2\xi_a^2\xi_b^2}\right)\Bigg], \tag{II–2.37}$$

$$v_i' = \frac{v_i^2-1}{\tau}\Bigg[2v_i\tau\mathcal{H} - \frac{k_i}{\xi_i} - \sum_{a,b} b^i_{ab}\frac{\bar{v}_{ab}}{2v_i}\frac{(\mu_a+\mu_b-\mu_i)}{\mu_i}\frac{\xi_i^2\ell^i_{ab}}{\xi_a^2\xi_b^2}\Bigg], \tag{II–2.38}$$

where ℓ^i_{ab} is the average length of segments of type i formed by the zipping/unzipping of string types a and b at conformal time τ, and μ_i is the tension of the $i^{\rm th}$ string type. b^i_{ab} are coefficients described below. All string tensions can be expressed in terms of the fundamental string tension $\mu_{\rm F}$, and in flat spacetime [291; 322; 379] are given by

$$\mu_i = \frac{\mu_{\rm F}}{g_{\rm s}}\sqrt{p_i^2 g_{\rm s}^2 + q_i^2}, \tag{II–2.39}$$

where p_i and q_i are the charges of string type i as listed in equation (II–2.32). The coefficients b^i_{ab} appearing in the velocity evolution equations (II–2.38) are related to energy conservation and allow for the energy saved from zipping interactions to be redistributed as kinetic energy of the new segment ($b^i_{ab} = d^i_{ab}$) [35] or radiated away ($b^i_{ab} = 0$) as in [360]. A more realistic model should have a specific radiation mechanism so that $0 < b^i_{ab} < d^i_{ab}$ such that some of the energy is redistributed whilst the rest is radiated away. However, for cosmic superstring networks (for which d_{ij} are much smaller than unity) this term has negligible impact on the string scaling densities and velocities [33; 293], so $b^i_{ab} = 0$ is used.

Once the velocities and correlation lengths of all string types in the network are obtained by solving (II–2.37 – II–2.38), their UETC can be calculated independently as laid out in Part II–2.1. Although $N > 3$ string types are needed in order to accurately construct the abundances of the dominant three lighter strings (in this case the seven string types in equation (II–2.32) are used), the resulting scaling densities of the higher charged states with $N > 3$ are strongly suppressed compared to the lighter F-, D- and FD-strings [33; 35; 360]. This

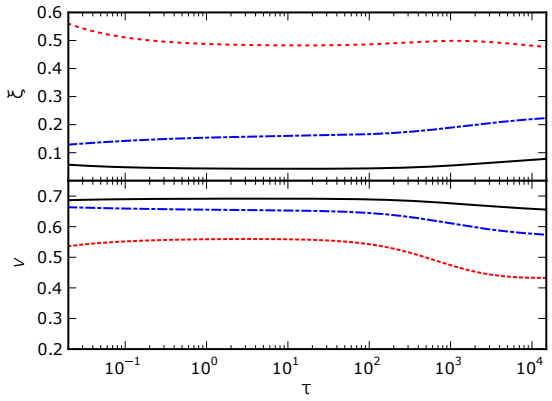


Figure II–2.8: The radiation and matter era evolution of the velocity v, and correlation length ξ, for the F-string in solid black, D-string in dot-dashed blue and FD-string in dotted red. These results are obtained when $g_{\rm s} = 0.3$, $w = 1$ and $c_{\rm s} = 0.23$.

allows only these first three states to be considered in the computation of CMB signatures through the UETC analytic method. The evolution of the network parameters for the three lightest strings can be seen in figure II–2.8 for $c_{\rm s} = 0.23$, $w = 1$ and $g_{\rm s} = 0.3$.

Once the UETC of each of the three lighter strings are calculated they can simply be summed to give the total string UETC, since the individual segments are uncorrelated in the USM. This can then be diagonalised and the eigenvectors and eigenmodes used as sources for finding the contribution from cosmic superstrings to the CMB anisotropy. The analytic UETC method reproduces the results of figure 4 in [293], including the shift in the location of the peak as $g_{\rm s}$ is varied. A slightly lower amplitude in the B-mode spectrum is found, and can be attributed to the extra factor of 2 in the vector modes that was present in `CMBACT3` (which [293] was based on) and has been corrected in `CMBACT4` [287]

2.3 Cosmic string constraints

Joint constraints on cosmic string network and ΛCDM parameters are obtained using a modified version of `COSMOMC`. To reduce computational time in the analysis two methods for deriving string network constraints have been tested. In the first method, the string C_ℓ are pre-calculated for a range of $c_{\rm r} = [0.1, 1]$ and $\alpha = [1, 10]$ at the *Planck* best fit values for the cosmological parameters, i.e. $\Omega_{\rm b}h^2$, $\Omega_{\rm c}h^2$ and H_0. These C_ℓ are read into `COSMOMC`, interpolated at the MCMC $c_{\rm r}$ and α values and then scaled by $(G\mu)^2$. This is an extremely efficient way for obtaining network constraints since only the ΛCDM C_ℓ need to be

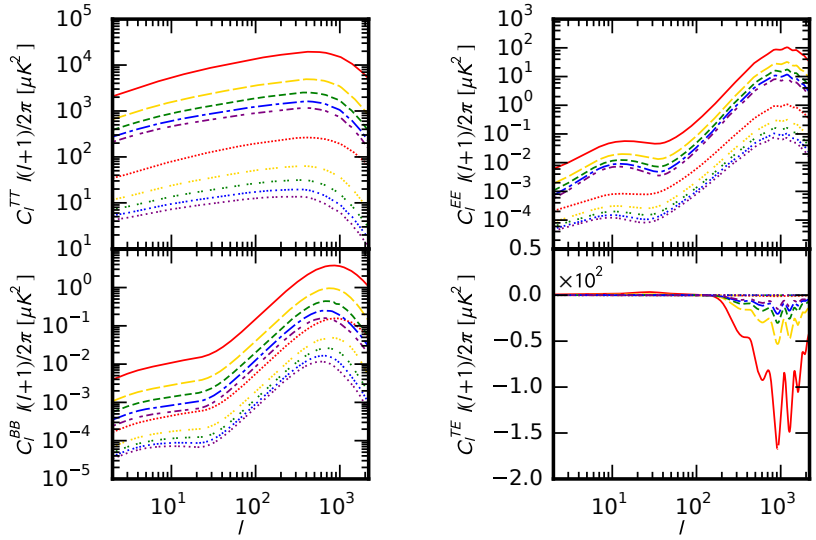


Figure II–2.9: The total C_ℓ (scalar+vector+tensor modes) for different values of c_r and α. The red solid lines show $c_\mathrm{r} = 0.1$ and through yellow (long-dashed), green (short-dashed), blue (dot-dashed) and purple (dot-dash-dotted) for $c_\mathrm{r} = 0.3$, $c_\mathrm{r} = 0.5$, $c_\mathrm{r} = 0.7$ and $c_\mathrm{r} = 0.9$. The upper (solid-patterned) lines indicate $\alpha = 10$ whilst the lower (dotted versions of the pattern) lines are for $\alpha = 1$. This is shown for C_ℓ^{TT} and C_ℓ^{EE} on the left and right of the top row, and C_ℓ^{BB} and C_ℓ^{TE} on the left and right of the bottom row.

calculated, while the interpolation takes very little time. The difference in the resulting string C_ℓ has been checked and when calculated at the upper and lower 3σ bounds in $\Omega_\mathrm{b}h^2$, $\Omega_\mathrm{c}h^2$ and H_0 is $\sim$0.5% in the temperature, E- and B-modes and no more than $\sim$ 10% in the TE cross-correlation. This uncertainty in the string C_ℓ is $\ll$ 1% of the total C_ℓ. The C_ℓ for different c_r and α are plotted in figure II–2.9. The different bands of colour indicate the value of c_r, solid red being the lowest ($c_\mathrm{r} = 0.1$) then progressing through long-dashed yellow, short-dashed green, dot-dashed blue and dot-dash-dotted purple in steps of 0.2, up to $c_\mathrm{r} = 0.9$. The upper (patterned) and lower (dot-patterned) edges of the bands indicate $\alpha = 10$ and $\alpha = 1$ respectively. From this it can be seen that the effect of α is to change the amplitude of the C_ℓ, with lower α also flattening the small ℓ features (as best seen in the upper right subplot and to a lesser extent in the lower left of figure II–2.9). Increasing c_r reduces the amplitude of the C_ℓ and, as best seen in the lower left subplot of figure II–2.9, shifts the main peak towards slightly smaller ℓ. In the second method, which is computationally expensive, the string and ΛCDM C_ℓ are simply calculated for each (network) parameter value and compared to CMB data.

The same process of pre-calculating string spectra can be done for cosmic superstring networks in the parameter ranges $c_{\rm s} = [0.1, 1]$, $g_{\rm s} = [0.01, 0.9]$ and $w = [0.001, 1]$. The superstring C_ℓ can be seen in figure II–2.10, where the same colours and patterns are used for the steps in $c_{\rm s}$ as in figure II–2.9. The bands indicate values of w, with $w = 10^{-3}$ corresponding to the solid-patterned lines and $w = 1$ to the dotted version of the same pattern. The rows indicate varying values of $g_{\rm s}$, with $g_{\rm s} = 0.01$, $g_{\rm s} = 0.1$ and $g_{\rm s} = 0.9$ for the top, middle and bottom rows respectively. The first point to notice is that the C_ℓ amplitudes at low $g_{\rm s}$ are much greater than those at large $g_{\rm s}$. For large $c_{\rm s}$ values there is less difference between the greatest and smallest values of w, especially at low $g_{\rm s}$, i.e. the purple dot-dash-dotted lines in the top row of figure II–2.10 overlap, but are well separated in the bottom row. This is because for large $c_{\rm s}$ the cross-interaction terms d_{ij}^k (which are less dependent on w than the self-interaction terms c_i) play a more important role in setting the scaling string number densities. For small values of $c_{\rm s}$, the c_i coefficients become smaller (while d_{ij}^k are unaffected) leading to small correlation lengths and so large string number densities. The C_ℓ amplitudes are then affected more strongly by c_i, giving rise to a stronger dependence on w.

The datasets used in the MCMC analysis come from the *Planck*2015 mission [18], in particular:

Planck2015 TT+lowP: This contains the 100-GHz, 143-GHz, and 217-GHz binned half-mission temperature autocorrelation (TT) frequency cross-spectra for $\ell = 30 - 2508$ with CMB-cleaned 353-GHz map, CO emission maps, and *Planck* catalogues for the masks and 545-GHz maps for the dust residual contamination template. It also uses the joint temperature and E-mode cross correlation (TE), E-mode autocorrelation (EE) and B-mode autocorrelation (BB) for $\ell = 2 - 29$ with EE and BB maps from the 70-GHz LFI full mission data and foreground contamination determined by 30-GHz LFI and 353-GHz HFI maps.

Planck2015 TT+Pol+lowP: This contains the same data as *Planck*2015 TT+ lowP but also uses the TE and EE cross-spectra for $\ell = 30 - 1996$.

Planck2015 TT+Pol+lowP+BKPlanck: This again contains all of the data used in *Planck*2015 TT+Pol+lowP but includes also the cross-frequency spectra between Background Imaging of Cosmic Extragalactic Polarization (BICEP2) and Keck maps at 150 GHz with *Planck* maps at 353 GHz including the B-mode spectra at multipoles $\ell \sim 50 - 250$.

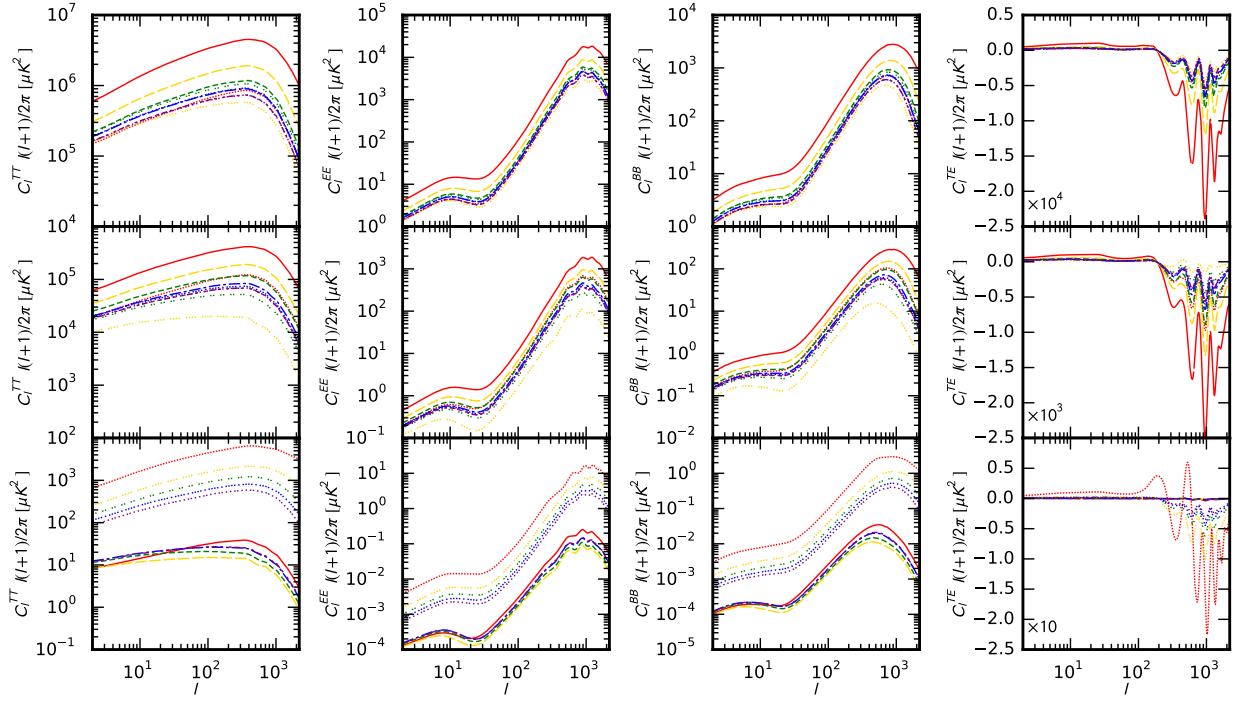


Figure II–2.10: The total C_ℓ obtained from cosmic superstrings for different values of $g_{\rm s}$, $c_{\rm s}$ and w. As with previous figures, the first column shows the temperature auto-correlation C_ℓ, the second the EE, third the BB and the last column shows the temperature, E-mode correlation C_ℓ. The three rows show $g_{\rm s}$ values of 10^{-2}, 10^{-1} and 0.9 from top to bottom. The colouring and patterning system is the same as in Fig. II–2.9 with red (solid), yellow (long-dashed), green (short-dashed), blue (dot-dashed) and purple (dot-dash-dotted) lines indicating the values of $c_{\rm s}$ from $c_{\rm s} = 0.1$ to 0.9 in steps of 0.2. The width of the band of similar colour and pattern indicates the upper and lower w values with the dotted-patterned lines defined by $w = 10^{-3}$ and the solid-patterned lines by $w = 1$.

The interpolation method is first considered, where the C_ℓ are pre-calculated on a grid in c_r and α (or in the case of cosmic superstring networks c_s, g_s and w), and then a spline interpolation used between grid values. The results obtained from this method are very quick and accurate due to the ability to use all 512 eigenmodes of the 512×512 grid for the UETC. The constraints on network parameters derived from this method are shown in figure II–2.11. $G\mu$ is implemented into the MCMC analysis through a logarithmic prior of $[-10, -5]$ such that $G\mu = 10^{[-10,-5]}$.

There is no significant difference in the constraints when using *Planck*2015 TT+lowP, or including EE and TE or both EE and TE and BB results. The upper 2σ value for the tension is $G\mu < 1.1 \times 10^{-7}$ for *Planck*2015 TT and is similarly $G\mu < 9.6 \times 10^{-8}$ and $G\mu < 8.9 \times 10^{-8}$ for *Planck*2015 TT+Pol+lowP and *Planck*2015 TT+Pol+lowP+BK*Planck*. These agree well with the $G\mu < 1.8 \times 10^{-7}$ and $G\mu < 1.3 \times 10^{-7}$ from the *Planck* cosmological parameters analysis [14]. The slightly tighter constraints obtained here are due to the amplitude of the C_ℓ not scaling with the value of L_f, i.e. the C_ℓ are larger when $L_\mathrm{f} = 1$ as assumed here, while previous results were obtained from `CMBACT` with $L_\mathrm{f} = 0.5$. There is little difference between using the *Planck* temperature data alone and including polarisation data as expected from [14]. As can be seen in the other two columns of figure II–2.11, c_r and α are not constrained. There is a slight preference for higher values of c_r and lower values of α since both of these lead to smaller C_ℓ. Features, such as the position of the main peak or the pronounced lower ℓ peak make very little difference to the overall constraints. There is a very slight correlation between $G\mu$ and c_r and anti-correlation between $G\mu$ and α, as expected from the C_ℓ seen in figure II–2.9. A combination of high α and low c_r is mildly disfavoured. Further, by comparing the constraints on $G\mu$ and c_r to their affect on the C_ℓ in figure II–2.9 there is a larger difference between changes at small c_r than changes at large c_r. For this reason we expect to see greater correlation between $G\mu$ and c_r on a logarithmic scale from values $c_\mathrm{r} \ll 1$ to $c_\mathrm{r} \approx 0.1$ than implied over the prior range used here.

Considering the direct calculation method, where the string spectra are calculated every time along with the C_ℓ from ΛCDM, the constraints are slightly weaker. This is because there is a pay-off between the resolution of the UETC and number of eigenmodes used in the reconstruction and the time spent computing the spectra. To efficiently calculate the constraints a grid resolution of 128×128 with 64 eigenmodes has been used. As can be seen in figure II–2.6 there

Figure II–2.11: 2σ likelihood contours for $G\mu$, $c_{\rm r}$ and α from the string C_ℓ interpolation and direct calculation methods. The orange line shows the constraints from *Planck*2015 TT+lowP, purple and blue lines are used for *Planck*2015 TT+Pol+lowP and *Planck*2015 TT+Pol+lowP+BK*Planck* respectively. The black dashed line shows the direct calculation constraints for *Planck*2015 TT+lowP.

is expected to be a reduction in power of about $10-20\%$ which means the value of $G\mu$ is allowed to be higher than when the high resolution, full reconstruction interpolation method is used. For *Planck*2015 TT+lowP this is $G\mu < 4.3\times10^{-7}$. The constraints on $c_{\rm r}$ and α also show a slight preference for lower $c_{\rm r}$ and larger α, as in the interpolation method.

For cosmic superstrings, $G\mu_{\rm F}$, $g_{\rm s}$ and w are marginalised over logarithmic priors, and $c_{\rm s}$ over a flat prior. Again all 512 eigenmodes of the 512×512 grid for the UETC are used. The likelihood contours obtained from the interpolation method can be found in figure II–2.12. It can be seen that w and $c_{\rm s}$ are almost flat (columns 3 and 4), again with larger values of $c_{\rm s}$ favoured as this leads to smaller amplitude C_ℓ. As the string density grows with decreasing $g_{\rm s}$, the constraints on $g_{\rm s}$ favour larger values, as seen in the second column. Note, however, that the model is not reliable for large values of $g_{\rm s}$ as the perturbative expansion starts to break down and the string interaction amplitudes used in c_i and d_{ij}^k have large uncertainties. Finally, the first column shows the constraints on the fundamental string tension $G\mu_{\rm F}$, which is much smaller than for ordinary cosmic strings. It can be seen that $G\mu_{\rm F} < 2.8\times10^{-8}$ for *Planck*2015 TT+lowP when marginalising over $g_{\rm s}$, $c_{\rm s}$ and w, and the same constraint for *Planck*2015 TT+Pol+lowP and *Planck*2015 TT+Pol+lowP+BK*Planck*.

Also figure II–2.12 shows the constraints when using the direct calculation method, where the string spectra are calculated at every step in the Markov chain. This is a much more intensive computation and so a lower resolution grid and fewer eigenmodes in the reconstruction had to be used. As for cosmic strings the optimal balance between computing time and accuracy suggested using a 128×128 grid with 64 eigenmodes. The constraints are thus slightly weaker, with the main result $G\mu_{\rm F} < 4.2\times10^{-8}$. The results from the two methods used here are in good agreement, justifying the use of the interpolation method, and showing that varying ΛCDM parameters within *Planck* priors has little effect on the string constraints.

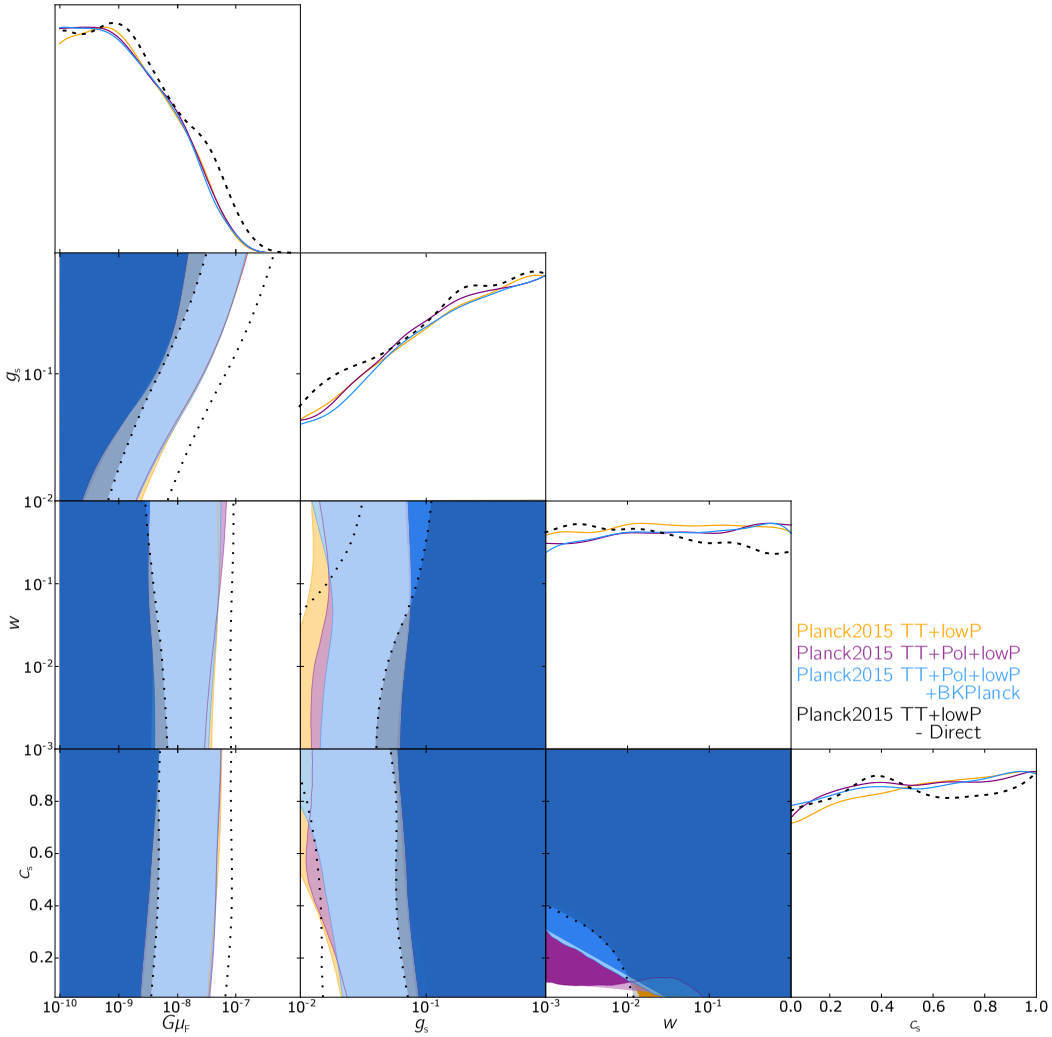


Figure II–2.12: 2σ likelihood obtained for $G\mu_{\rm F}$, $g_{\rm s}$, w and $c_{\rm s}$. The black dashed line (and black shading) shows the constraints from *Planck*2015 TT+lowP direct calculation method and orange, purple and blue lines are the *Planck*2015 TT+lowP, *Planck*2015 TT+Pol+lowP and *Planck*2015 TT+Pol+lowP+BK*Planck* constraints using the interpolation method.

Chapter 3

Discussion

Currently, there are two main approaches to the detection of cosmic strings. Firstly, since they actively generate scalar, vector and tensor perturbations they lead to signatures in the temperature, polarisation, and non-Gaussian spectra of the CMB. Secondly, a cosmic string network will emit gravitational waves, primarily from loop decay. This leads to a stochastic background which can be constrained using pulsar timing, laser interferometry experiments such as the Laser Interferometer Gravitational-Wave Observatory (LIGO) and extendend Laser Interferometer Space Antenna (eLISA), and also the CMB [338]. A transient gravitational wave signal is also expected from cusps and kinks in the network [98]. The latter class of tests has the potential to provide even stronger constraints on the string tension $G\mu$, but there are large uncertainties in the loop size, which is fixed by gravitational back-reaction. Model dependence on gravitational waves from cosmic strings further makes it difficult to determine signatures, for example, whilst Nambu-Goto strings decay into loops, Abelian-Higgs strings primarily decay into particles [98; 115; 273]. It is therefore important to use a variety of complementary observational probes.

The first class of tests also suffer from uncertainties, but these are less significant. The string UETC can be obtained from simulations and used as source functions in CMB codes, but simulations are numerically expensive and suffer from issues in dynamical range. An alternative approach is to model the string network as an ensemble of segments using the USM. Crucially, although the USM provides a simplified picture of the network, it is able to match simulations by adjusting the free parameters of the model, namely the correlation length, RMS velocity and string wiggliness.

In this Part of the Thesis, previous work on string power spectra from the USM

has been significantly improved and extended.

1. The UETC has been analytically solved for an evolving string network, compared with previous work where it was restricted to constant network parameters. The UETC itself can be computed in under a minute. For the CMB power spectrum, although the time taken is increased due to tracking a larger number of Fourier modes, on a 3.1 GHz Intel Xeon central processing unit (CPU) with 8 threads, the code runs in ~ 60 minutes. For comparison, around 2000 network realisations are required for `CMBACT4` to achieve the same accuracy and since this code is serial, the computation time is ~ 30 hours.

2. The formalism has been extended to cosmic superstring networks with multiple string types and different network parameters. Here the UETC can be computed for each string type and added, since the segments are assumed to be uncorrelated. The UETC calculation is much quicker than the CMB line-of-sight integration, so the total computation time is not significantly increased over the single string case.

3. For the first time marginalisation over the string network parameters when fitting to *Planck*2015 and joint *Planck*-BICEP2 data has been achieved. The data is consistent with no strings for both the single and multi-string case. Since other network parameters are unconstrained when the tension is very small, it is only possible to present joint constraints on these with $G\mu$. In the superstring case, for example, the constraint on the string coupling $g_{\rm s}$ is degenerate with $G\mu_{\rm F}$.

There are several possibilities to explore in future work. Firstly, there are various ways in which the USM could be improved. Superstring networks contain Y-type junctions, but in the present formulation these only impact the evolution of the network parameters. Since junctions are relatively rare in the limit of large and small coupling, the USM is expected to provide a sufficient description. However, in some regimes the energy density of the network may not be dominated by a single string type, and junctions may become important. In this case the USM could be modified to include a correlation between segments. A further improvement is the inclusion of loops. The decay of string segments in the USM should mimic the energy loss in loops, but it is possible these may lead to additional interesting signatures.

Given that *Planck* has largely exhausted the available signal in the temperature data, future string constraints from the CMB will be driven by polarisation

and non-Gaussianity. The non-Gaussian signal from post-recombination simulations has been used to obtain constraints on $G\mu$ [13], and attempts have been made to compute the bispectrum analytically using a Gaussian model for the string correlators [296]. It is also possible to compute the non-Gaussian signal using the USM which will, by design, include physics from recombination and along the line-of-sight. This has already been demonstrated for the CMB bispectrum [137] by performing many realisations of the network. It is possible to employ a similar analytic method used in this work to compute the string bispectrum and trispectrum, which is expected to be significantly faster [82].

The detection of gravitational waves by LIGO is particularly exciting for strings, and the next generation of ground and space based experiments can potentially provide much stronger limits than those from the CMB. However, these limits strongly depend on modelling, for example, the loop, kink and cusp distribution. Further work is needed to understand these and until then, the CMB will continue to be an important tool in the search for strings.

Appendix A

Integral identities and analytic coefficients

$$
\begin{aligned}
I_1(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin\theta \sec^2\theta \cos(x\cos\theta) J_0(\varrho\sin\theta) \\
&= \sum_{c=0}^{\infty} \frac{1}{c!}\frac{\varrho}{(2c-1)}\Big(-\frac{x^2}{2\varrho}\Big)^2 j_{c-1}(\varrho)
\end{aligned}
\tag{II–A.1}
$$

$$
\begin{aligned}
I_2(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin\theta \cos(x\cos\theta) J_0(\varrho\sin\theta) \\
&= \left(\frac{\sin\sqrt{\varrho^2+x^2}}{\sqrt{\varrho^2+x^2}}\right)
\end{aligned}
\tag{II–A.2}
$$

$$
\begin{aligned}
I_3(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin^3\theta \cos(x\cos\theta) J_0(\varrho\sin\theta) \\
&= \left[1+\frac{\partial^2}{\partial x^2}\right]\left(\frac{\sin\sqrt{\varrho^2+x^2}}{\sqrt{\varrho^2+x^2}}\right)
\end{aligned}
\tag{II–A.3}
$$

$$
\begin{aligned}
I_4(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin\theta \sec^2\theta \cos(x\cos\theta) \frac{J_1(\varrho\sin\theta)}{\varrho\sin\theta} \\
&= \frac{\cos x}{\varrho^2} - \sum_{c=0}^{\infty}\frac{1}{c!}\frac{1}{(2c-1)}\Big(-\frac{x^2}{2\varrho}\Big)^2 j_{c-2}(\varrho)
\end{aligned}
\tag{II–A.4}
$$

$$
\begin{aligned}
I_5(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin\theta \cos(x\cos\theta) \frac{J_1(\varrho\sin\theta)}{\varrho\sin\theta} \\
&= \frac{1}{\varrho^2}\left(\cos x - \cos\sqrt{\varrho^2+x^2}\right)
\end{aligned}
\tag{II–A.5}
$$

$$
\begin{aligned}
I_6(x,\varrho) &= \frac{1}{2}\int_0^\pi d\theta \sin^3\theta \cos(x\cos\theta) \frac{J_1(\varrho\sin\theta)}{\varrho\sin\theta} \\
&= -\frac{1}{\varrho^2+x^2}\left[1+\frac{1}{x}\frac{\partial}{\partial x}\right]\left(\cos\sqrt{\varrho^2+x^2}\right)
\end{aligned}
\tag{II–A.6}
$$

	$\langle\Theta_{00}(\tau_1,k)\Theta_{00}(\tau_2,k)\rangle$	$\langle\Theta^{\mathrm{S}}(\tau_1,k)\Theta^{\mathrm{S}}(\tau_2,k)\rangle$	$\langle\Theta^{\mathrm{V}}(\tau_1,k)\Theta^{\mathrm{V}}(\tau_2,k)\rangle$	$\langle\Theta^{\mathrm{T}}(\tau_1,k)\Theta^{\mathrm{T}}(\tau_2,k)\rangle$	$\langle\Theta^{\mathrm{S}}_{00}(\tau_1,k)\Theta^{\mathrm{S}}_{00}(\tau_2,k)\rangle$
a_1	$2\alpha^2$	$\frac{1}{2\alpha^2}$	0	$\frac{1}{4\alpha^2}$	1
b_1	0	$1-\frac{1}{2\alpha^2}$	0	$-\frac{1}{4\alpha^2}$	$-\frac{1}{2}+\alpha^2$
c_1	0	$\frac{1}{2\alpha^2}-2+2\alpha^2-\frac{27\alpha^2}{2\varrho^2}$	$\frac{3\alpha^2}{\varrho^2}$	$\frac{1}{4\alpha^2}-\frac{3\alpha^2}{4\varrho^2}$	0
a_2	0	$\frac{3}{2\alpha^2}$	0	$-\frac{1}{4\alpha^2}$	-3
b_2	0	$-\frac{3}{2\alpha^2}$	0	$\frac{1}{4\alpha^2}$	$\frac{3}{2}-\frac{3\alpha^2}{2}$
c_2	0	$\frac{3}{2\alpha^2}-\frac{3\alpha^2}{2}+\frac{27\alpha^2}{2\varrho^2}$	$-\frac{3\alpha^2}{\varrho^2}$	$-\frac{1}{4\alpha^2}+\frac{3\alpha^2}{4\varrho^2}+\frac{\alpha^2}{4}$	0
a_3	0	$-\frac{9}{2\alpha^2}$	$\frac{1}{\alpha^2}$	$-\frac{1}{4\alpha^2}$	0
b_3	0	$\frac{9}{2\alpha^2}-\frac{9}{2}$	$1-\frac{1}{\alpha^2}$	$\frac{1}{4\alpha^2}-\frac{1}{4}$	0
c_3	0	$-\frac{9}{2\alpha^2}+9-\frac{9\alpha^2}{2}$	$\frac{1}{\alpha^2}-2+\alpha^2$	$-\frac{1}{4\alpha^2}+\frac{1}{2}-\frac{\alpha^2}{4}$	0
a_4	0	0	0	0	0
b_4	0	$-\frac{3}{2}$	0	$\frac{1}{4}$	$-\frac{3\alpha^2}{2}$
c_4	0	$3-6\alpha^2+\frac{27\alpha^2}{\varrho^2}$	$\alpha^2-\frac{6\alpha^2}{\varrho^2}$	$-\frac{1}{2}+\frac{3\alpha^2}{2\varrho^2}$	0
a_5	0	0	0	0	0
b_5	0	$\frac{3}{2}$	0	$-\frac{1}{4}$	$\frac{3\alpha^2}{2}$
c_5	0	$-3+6\alpha^2-\frac{27\alpha^2}{\varrho^2}$	$-\alpha^2+\frac{6\alpha^2}{\varrho^2}$	$\frac{1}{2}-\frac{3\alpha^2}{2\varrho^2}$	0
a_6	0	0	0	0	0
b_6	0	$\frac{9}{2}$	-1	$\frac{1}{4}$	0
c_6	0	$-9+9\alpha^2$	$2-2\alpha^2$	$-\frac{1}{2}+\frac{\alpha^2}{2}$	0

Table II–A.1: Coefficients for the amplitude equations given by $A_i = a_i + b_i(v(\tau_1)^2 + v(\tau_2)^2) + c_i v(\tau_1)^2 v(\tau_2)^2$.

	$\langle\Theta_{00}(\tau_1,k)\Theta_{00}(\tau_2,k)\rangle$	$\langle\Theta^{\rm S}(\tau_1,k)\Theta^{\rm S}(\tau_2,k)\rangle$	$\langle\Theta^{\rm V}(\tau_1,k)\Theta^{\rm V}(\tau_2,k)\rangle$	$\langle\Theta^{\rm T}(\tau_1,k)\Theta^{\rm T}(\tau_2,k)\rangle$	$\langle\Theta^{\rm S}_{00}(\tau_1,k)\Theta^{\rm S}_{00}(\tau_2,k)\rangle$
B	$\alpha^2 x_1 x_2$	$\frac{x_1x_2}{5\alpha^2}X$	$\frac{x_1x_2}{15\alpha^2}X$	$\frac{x_1x_2}{15\alpha^2}X$	0
C	Z^a	$YZ^a+\alpha^2 v(\tau)^4 Z^b$	$YZ^a+\alpha^2 v(\tau)^4 Z^b$	$YZ^a+\alpha^2 v(\tau)^4 Z^b$	$[Y-v(\tau)^4(1-\alpha^2)]Z^a$
z_1^a	$-2\alpha^2$	$-\frac{2}{\alpha^2}$	$\frac{2}{3\alpha^2}$	$\frac{-2}{3\alpha^2}$	-4
z_2^a	$2\alpha^2$	$\frac{1}{2\alpha^2}-\frac{9}{\alpha^2x^2}$	$\frac{2}{\alpha^2x^2}$	$\frac{1}{4\alpha^2}-\frac{1}{2\alpha^2x^2}$	1
z_3^a	0	$-\frac{3}{2\alpha^2x}+\frac{9}{\alpha^2x^3}$	$-\frac{2}{\alpha^2x^3}$	$\frac{1}{4\alpha^2x}+\frac{1}{2\alpha^2x^2}$	$\frac{3}{x}$
z_4^a	$2\alpha^2$	$\frac{1}{2}$	0	$\frac{1}{4\alpha^2}$	1
z_1^b	0	-2	0	0	0
z_2^b	0	$\frac{11}{16}-\frac{27}{8x^2}$	$\frac{1}{8}+\frac{3}{4x^2}$	$\frac{3}{32}-\frac{3}{16x^2}$	0
z_3^b	0	$\frac{3}{16x}+\frac{27}{8x^3}$	$\frac{1}{8x}-\frac{3}{4x^3}$	$-\frac{5}{32x}+\frac{3}{16x^3}$	0
z_4^b	0	$\frac{11}{16}$	$\frac{1}{8}$	$\frac{3}{32}$	0

Table II–A.2: Coefficients for the small x approximation and the ETC, expressed in terms of the functions $X = \left[1-(v(\tau_1)^2+v(\tau_2)^2)\left(1-\frac{\alpha^2}{2}\right)+v(\tau_1)^2v(\tau_2)^2(1-\alpha^2+\alpha^4)\right]$, $Y=\left[1-v(\tau)^2(2-\alpha^2)+v(\tau)^4(1-\alpha^2)\right]$ and $Z^j = z_1^j + z_2^j\cos x + z_3^j \sin x + z_4^j x\mathrm{Si}[x]$.

Part III

Statistical cosmology

There is a huge amount of data which needs to be understood to quantify parameters in models of cosmology. As mentioned in Part I, a basic, six-parameter cosmological model, ΛCDM, is able to fit the observed data exceptionally well, from both measurements of the CMB and from various LSS probes [14; 17; 57; 199]. The *yearning for learning* about parameters values, in a given model, has kickstarted an in depth look at statistical quantification of cosmological data [178; 358]. Advanced techniques such as MCMC analysis and importance sampling allows constraints to be placed on high dimensional cosmological parameter distributions. In particular, the use of MCMC code `COSMOMC` [234] built on the fast Einstein-Boltzmann cosmology solver `CAMB` [237] allow for various data sets to provide constraints on cosmology by creating samples probing the parameter distributions of cosmological parameters.

It is through disagreement between predicted model parameters and observed data that changes to a particular model are considered. The establishment of the ΛCDM model itself occurred through these channels. Although the clinching piece of evidence was the detection of cosmic acceleration using type-Ia supernovae [285; 302], there had been clear indications that a universe with critical matter density did not fit the data. Early in the process of constraining cosmological parameters it was seen that the RMS perturbation in spheres of radius $8h^{-1}\,\mathrm{Mpc}$, σ_8, for the COBE normalised critical matter density models was $\sigma_8 \approx 1.5$, whereas observations from a range of indicators suggested that it was in the range $0.7 - 0.9$ favouring $\Omega_\mathrm{m} < 1$ [125; 168]. The shape of the matter power spectrum was also in conflict with $\Omega_\mathrm{m} = 1$ [73; 122].

As the precision of cosmological parameter estimation increased, it was noticed that the parameter values of Ω_m and σ_8 were in contention when constrained by different data [5; 12; 45; 57; 112; 138; 156; 183; 222; 245; 300; 310; 382]. The implication was that constraints from probes of LSS implied there was too much small scale structure when compared to the constraints from measurements of the CMB. In particular, it was shown in [45; 156; 382] that both SZ cluster counts and lensing, from the CMB and from cosmic shear, were in conflict with CMB measurements and that a neutrino component - which could be from active or sterile neutrinos - could be used to reconcile these measurements. In turn, this built on the earlier suggestion in [11] that tension between the CMB measurements and SZ cluster counts could be accounted for in this way. The fact that the SZ cluster counts and the lensing data are compatible with each other strengthens the two $\sim 2\sigma$ discrepancies into a statistically improbable discrepancy of $\sim 4\sigma$. This reconciliation of measurements of large and small scales

was at the expense of less good fit to the CMB data - the two being seen to be in conflict at the level of 2.8σ [45].

Neutrino masses are an obvious way to explain a dearth of power on small scales. Particle physics oscillation experiments are sensitive to the square differences between the neutrino masses and cosmology is mostly sensitive to the sum of the masses, $\sum m_\nu$. If the neutrinos are sterile it is possible that the best-fit model can be made more compatible with direct measurements of the Hubble constant from low-redshift standard candles such as Cepheids [301]. The preference for massive neutrinos reported in [45; 156; 382] is a result of a global fit with an extended cosmological model. Clearly any systematic error in the data, or its interpretation, could lead to a false detection and therefore it needs to be treated with caution. Moreover, there are other extensions to the standard model that could lead to a similar result.

Bayesian statistics has plenty of tools which can help distinguish between models given posterior distributions of parameters [178] (and see Part I–3.4). Although model comparison is routine, it is less common to test the significance of data sets within a given model. In Part III–1, techniques for comparing two high dimensional posterior distributions given a model but two different data sets is discussed. There are several accepted techniques which can lead to misleading interpretations and so two new measures are introduced to better understand these differences.

In Part III–2 the probability distributions of ΛCDM parameters obtained by measurements of the CMB and a range of LSS probes are compared. The tension between the parameter constraints are discussed using the various measures introduced in Part III–1. The results themselves can be interpreted differently depending on the belief in well understood cosmological measurements, i.e. the size of the sound horizon $\Theta_{\rm MC}$ and the scalar spectral index $n_{\rm s}$. If these are constrained freely by the LSS measurements then the two sets of parameter constraints are, mostly, in good agreement. But, if well believed tight priors on $\Theta_{\rm MC}$ and $n_{\rm s}$ are used, the tension between parameter constraints increases greatly due to correlation between parameter distributions. To alleviate the discordance arriving in this case, it is useful to discuss extensions to ΛCDM. A non-exhaustive range of alternative explanations for the tension, including added neutrino content, optical depth assumptions inferred from WMAP polarisation, and modifications to the primordial power spectrum are present in Part III–2.4.

It should be noted that work in this Part comes from two different analyses. The first, carried out in 2013-2014, had Gaussian priors on $\Theta_{\rm MC}$ and $n_{\rm s}$ which came from the *Planck*2013 analysis. The second set of results come from 2015-2016 and do not contain these tight priors. Since the same procedure was, mostly, carried out for both studies, they are presented side by side.

Chapter 1

Quantifying discordance

The probability distribution $P(\theta)$ of the five relevant ΛCDM parameters excluding τ which is only constrained by the CMB, is a complicated 5D, not-necessarily Gaussian, function. These parameters were introduced in Part I–3.3. When constraining the parameters using the CMB only, one distribution $P(\theta|\text{CMB}, \Lambda\text{CDM})$ is found and a second, supposedly similar, distribution $P(\theta|\text{LSS}, \Lambda\text{CDM})$, can be derived from constraints using LSS. Since each of these distributions are difficult to quantify in a simple way, any comparison between them is also complicated. A number of different measures are used to give a simple, generally "single-numbered", quantification of any differences [61; 190], where the measures introduced in [247; 323; 367] are particularly used in cosmology and astronomy. The way each of these measures are interpreted can lead to confusing statements about any discordance and so a thorough discussion of a few of the major methods is laid out here. In the next chapter, each of the methods mentioned here will be used to quantify the tension between parameter constraints using CMB data and LSS data. Detailed descriptions of each method, using some simple distributions, can be found in Appendix III–B in order to help guide the reader.

1.1 Methods to quantify discordance

Consider the posterior distributions $P_1 \equiv P(\theta|D_1, \mathcal{M})$ and $P_2 \equiv P(\theta|D_2, \mathcal{M})$ for data sets D_1 and D_2, respectively, parameters, θ, of a model $\mathcal{M}$.

1. Bhattacharyya distance The Bhattacharyya distance [61] compares the probability distributions from each model at a given parameter value

$$B = \int d\theta \sqrt{P_1 P_2}. \tag{III–1.1}$$

$B = 1$ indicates two identical distributions whilst $B \gtrsim 0$ for disparate distributions with values in between indicating the level of tension. If one of the distributions is particularly broad compared to the other then this will give a low Bhattacharyya distance value meaning the distributions are distinctly different. This is true even if the peaks of the distributions are identical. The Bhattacharyya distance is not used in a cosmological context since the variance of the posterior distribution given LSS data is often much wider than when using measurements of the CMB. It is, however, easy to understand and aids in comprehension of comparisons between posterior distributions.

2. Overlap coefficient The overlap coefficient [190] works in a similar way to the Bhattacharyya distance. In this case the quantity obtained is given by

$$O = \int d\theta \mathrm{Min}[P_1, P_2]. \quad \text{(III–1.2)}$$

As with B, two identical distributions have $O = 1$ and non-overlapping distributions have $O = 0$. The scale of difference between $0 < O < 1$ is not the same as the Bhattacharyya distance, with the overlap coefficient taking lower values for the same pair of differing distributions. Again broader distributions are indicated as being in *tension*, even with identical distribution peaks. This is also not often used for cosmological comparison.

3. Difference vector This measure, presented in detail here and inspired by the two sample T-test [299], involves calculating the difference between the parameter ranges from the first and second probability distributions and creating a new probability distribution from the difference vector

$$P(\delta\theta|D_1, D_2, \mathcal{M}) = \int d\theta' P_1(\theta') P_2(\theta' - \delta\theta). \quad \text{(III–1.3)}$$

Here $\delta\theta = \theta_1 - \theta_2$, where θ_1 and θ_2 are the allowed values of the parameters from the distributions from data set D_1 and data set D_2, thus $P_2(\theta_1 - \delta\theta) \equiv P_2(\theta_2)$. This convolution effectively shifts the mean of the new distribution to the difference in the means of the original two distributions, $\mu_{\delta\theta} = \mu_{\theta_1} - \mu_{\theta_2}$, with parameters spanning a range from $\mu_{\delta\theta} - \mathrm{Min}[\theta_1, \theta_2]$ to $\mu_{\delta\theta} + \mathrm{Max}[\theta_1, \theta_2]$. For convenience $P(\delta\theta|D_1, D_2, \mathcal{M})$ is denoted $P(\delta\theta)$. A quantification of the disagreement between the distributions is obtained by integrating this new distribution within the isocontour formed by the value of the probability distribution function at $\delta\theta = \mathbf{0}$,

$$C = \int_A d\delta\theta P(\delta\theta), \quad \text{(III–1.4)}$$

where

$$A = \{\delta\theta \,|P(\delta\theta) > P(\mathbf{0})\} \tag{III–1.5}$$

For the analysis in Part III–2 when using *Planck*2013+WP+BAO (defined in Part III–2.1) for the CMB data, samples are taken from MCMC chains and analysed, giving means for each parameter and a covariance matrix for each distribution. The covariance matrices are then combined using the law of total covariance [309]. This combined covariance is used to form a multivariate Gaussian distribution centred at the difference in the means of the parameters obtained from the `COSMOMC` analysis of the MCMC chains. When using *Planck*2015+Pol+BAO (also defined in Part III–2.1) for the CMB data in Part III–2 the difference between the samples in the chains is used directly to form the probability distribution. This means that any non-Gaussianity of the distributions is taken into account.

As a single unit measure this does a good job of indicating disagreements between distributions. It can be interpreted easily since C is a measure of the fraction of samples within a bounded area. This area is arbitrary and choosing $\delta\theta = \mathbf{0}$ (as in equation (III–1.5)) is not essential. Of course, the measure cannot fully describe the complexity of both of the entire probability distribution functions P_1 and P_2. Using more parameters can help give greater understanding.

4. Integration between intervals Using two numbers to quantify the similarities and differences between probability distributions *can* provide more information. By integrating each of the probability distributions within a given interval of the other distribution, the total level of agreement can be quantified. The two useful numbers here are

$$I_1 = \int_{A_2} d\theta P_1 \tag{III–1.6}$$

$$I_2 = \int_{A_1} d\theta P_2, \tag{III–1.7}$$

where

$$A_i = \left\{\theta \,\middle|\, \int d\theta P_i = 0.997\right\}. \tag{III–1.8}$$

I_1 is obtained by integrating the probability distribution P_1 within the isocontour of the probability distribution P_2 which would contain 99.7% of the samples drawn from it. I_2 is obtained in exactly the same way, exchanging the probability distribution P_1 for P_2. This measure is particularly useful since I_1 and I_2 can be directly related to samples obtained via MCMC analysis. The limit chosen

for the integration interval is arbitrary. If the interval is chosen to measure the amount of P_1 within the isocontour which contains 68.4% of P_2 then, if $I_1 = 0$, the *tension* could be interpreted as being greater than 1σ. We have chosen to consider isocontours containing 99.7% of the samples from each distribution. If $I_1 = 0$ when integrated within the bounds containing 99.7% of the samples drawn from P_2 then P_1 would be considered to be in $> 3\sigma$ tension with P_2. Although computationally intensive, this method can be used to quantify an exact tension by increasing the integration limits of one of the distributions until the integral of the other distribution was no longer zero. This procedure is not performed in Part III–2.3 due to computational resources.

5. Surprise Another method which compares one distribution to another giving two measures is that used in [323]. Here the relative entropy is found when P_2 is used as an update to P_1 and is given by

$$D(P_1||P_2) = \int d\theta P_2 \log \frac{P_2}{P_1}. \quad \text{(III–1.9)}$$

An expected relative entropy can be found using

$$\langle D \rangle = \int dP_2 \int d\theta P_1 P_2 D(P_1||P_2). \quad \text{(III–1.10)}$$

By comparing the difference of the relative entropy to the expected relative entropy a quantity (which is named *surprise* in [323]) can be calculated

$$S = D(P_1||P_2) - \langle D \rangle. \quad \text{(III–1.11)}$$

Using a combination of $D(P_1||P_2)$ and S a quantification of information gain due to different distributions can be found. S should be close to zero for data sets which are similar and can be positive or negative. A positive *suprise* indicates that the distribution used to update the original is more different than expected. A negative *suprise* is obtained when the updating distribution is in more agreement than expected with the original distribution. This technique is particularly useful when comparing the amount of *suprise* for a given expected relative entropy. The results of which can be quoted as a p-value and interpreted as how likely one distribution is to be in agreement with the other.

6. Quantification of Bayesian evidence Other measures that have previously been discussed generally involve comparisons of Bayesian evidences. The most simple and commonly used was introduced in [247]. This is given by

$$R = \frac{p(D_1, D_2)}{p(D_1)p(D_2)}, \quad \text{(III–1.12)}$$

where $p(D_i)$ is the evidence given data D_i,

$$p(D_i) = \int d\theta P_i p(\theta), \tag{III–1.13}$$

where $p(\theta)$ is the prior on the parameter θ and the index $i = 1, 2$ denotes which data set is used. The numerator of equation (III–1.12) is given by

$$p(D_1, D_2) = \int d\theta P_1 P_2 p(\theta). \tag{III–1.14}$$

This is related quite closely to the Bhattacharyya distance. R is the ratio of the evidence given both data sets, to the evidence of each data set. The prior assumptions of the parameter must be specified and taken into account. Using $\log R$, the results can be interpreted on the Jefferys scale with $\log R > 0$ indicating agreement and $\log R < 0$ indicating disagreement to some degree. This, as for the Bhattacharyya distance and the overlap coefficient methods, reveals a degeneracy between shifts in the peaks of distributions and broadening of the variances of distributions. The numbers from $\log R$ are dependent on the choice of priors. As long as the prior is stated along with analysis then the results can be recreated and interpreted by the individual.

7. Shifted probability distribution Another measure, used in [366], shifts one distribution (in a similar way to the difference vector method) so that the maxima of the two distributions coincide is then found. The ratio to the joint probability distribution

$$T = \frac{p(D_1, D_2)_{\text{shifted}}}{p(D_1, D_2)}. \tag{III–1.15}$$

Identical distributions have $\log T = 0$ and $\log T > 0$ indicates deviations from similarity. The values of $\log T$ do not directly map to a statistical significance or a p-value. Also, $\log T$ can be expected to be twice as large when the dimensionality of the problem increases by two. This can either be corrected or taken into consideration when interpreting the result.

Each of the measures described in this section indicate, to some degree, whether or not two distributions agree or disagree with each other. They do not each give the same emphasis as to *where* tension arises.

- The Bhattacharyya distance, overlap coefficient and quantification of Bayesian evidence give disagreements arising from broadening of one distribution in comparison to another. The difference vector, shifted probability distribution, integration between intervals and *suprise* methods take this broadening into account.

- The Bhattacharyya distance and overlap coefficient have results which are difficult to interpret and do not map to any useful scales.
- The quantification of Bayesian evidence and shifted probability distribution methods are prior dependent and, out of the two, only $\log R$ can be interpreted on the Jeffreys scale.
- The *suprise* gives a variety of quantifications which can be mapped to two p-values, thus quantifying the amount of disagreement when either distribution is used to update the other.
- The difference vector relates the fraction of samples within an arbitrary boundary formed by the samples away from the difference in the means. It does not capture all the information, but can be quoted as a single number by mapping C onto the interval of the 1D Gaussian. Due to its construction, the value of C matches the expected results when comparing 2D likelihood contours, but extends to higher dimensions.
- Integration between intervals is more powerful than using C for observing differences and it is easy to understand each integral individually. However, the combination needs to be taken into account to truly describe how much tension is present between distributions. This can lead to some confusion when considering a broad distribution compared with a tight one.

In Part III–2.3, the difference vector measure (*3*) will be used for comparison of the constraints on ΛCDM parameters derived from the CMB and individual LSS probes. This represents an update of [44] (also considered in Part III–2.3) on the basis of more recent data.

Chapter 2

Cosmological parameter constraints

In this chapter a range of CMB and LSS data sets are used for the comparison of the posterior distribution of cosmological parameters in ΛCDM and its extensions. Since probability distributions are complex and multi-dimensional, care needs to be taken when making histograms from MCMC chain samples. These distributions can often be sparsely sampled in important overlapping regions. For measures *1*, *2* and *4*-*7* from Part III–1 the histogram of the distributions considered in this chapter are made for a number of different bins and both with and without Gaussian smoothing. The results quoted are the consensus values from this range of tests (which are all quite similar). For measure *3*, the number of samples from the chains is much greater since there are $N_{\rm CMB} \times N_{\rm LSS}$ differences, where $N_{\rm CMB}$ is the number of samples from the CMB chains and $N_{\rm LSS}$ is the number of samples from the LSS chains. The histogram is made for a range of bins and with and without Gaussian smoothing to check that the results are robust. The data used is presented below.

2.1 Cosmic microwave background

The temperature anisotropies and polarisation of photons from the CMB have been measured to an extremely high resolution over the largest possible scales by both *Planck* [7; 14] and WMAP [176] with smaller scale analysis performed by the Atacama Cosmology Telescope (ACT) [100] and the South Pole Telescope (SPT) [298]. Several combinations of CMB data will be used.

WMAP+highL+BAO: The 9 year full mission temperature measurements made by WMAP [176] over the multipole range $\ell \approx 2 - 800$ complemented with

higher resolution measurements made by ACT [100] and SPT [298] and combined with BAO discussed below:

Planck2013+WP+BAO: Measurements of the temperature anisotropy power spectrum made by *Planck* that have been extensively used for cosmological parameter analysis [9]. These results cover the multipole range $\ell \approx 2 - 2500$. This is implemented by using the standard likelihood [8] and uses the measurements of the polarisation and temperature-polarisation cross-correlation power spectra from WMAP 9 year data [177]. BAO are also included and described below.

Planck2015+Pol+BAO: The updated results from the *Planck*2015 analysis. This includes the temperature TT, EE and TE spectra from *Planck* HFI for $29 < \ell < 2509$ and TT, TE, EE and BB spectra from *Planck* LFI for $2 < \ell < 29$ [18]. This is combined with the measure of the BAO peak described below.

The ratio of the sound horizon at the drag epoch, $r_{\rm s}(z_{\rm d})$, to the volume-averaged distance $D_{\rm V}(z_{\rm eff})$, can be constrained using BAO ([58; 142])

$$\frac{D_{\rm V}(z_{\rm eff})}{r_{\rm s}(z_{\rm d})} = \frac{\left(\alpha_\perp^2 \alpha_\parallel [(1+z_{\rm eff}) D_{\rm A}^{\rm fid}(z_{\rm eff})]^2 \frac{z_{\rm eff}}{H^{\rm fid}(z_{\rm eff})}\right)^{1/3}}{r_{\rm s}^{\rm fid}(z_{\rm d})}. \quad \text{(III–2.1)}$$

$D_A^{\rm fid}(z_{\rm eff})$, $H^{\rm fid}(z_{\rm eff})$ and $r_{\rm s}^{\rm fid}$ are the fiducial values of the angular diameter distance, Hubble function and sound horizon at the drag epoch for a given cosmology. The scaling factors along the line-of-sight and perpendicular to it are, respectively ([58])

$$\alpha_\parallel = \frac{H^{\rm fid}(z_{\rm eff}) r_{\rm s}^{\rm fid}(z_{\rm d})}{H(z_{\rm eff}) r_{\rm s}(z_{\rm d})}, \quad \text{(III–2.2)}$$

and

$$\alpha_\perp = \frac{D_{\rm A}(z_{\rm eff}) r_{\rm s}^{\rm fid}(z_{\rm d})}{D_{\rm A}^{\rm fid}(z_{\rm eff}) r_{\rm s}(z_{\rm d})}. \quad \text{(III–2.3)}$$

Both WMAP+highL+BAO and *Planck*2013+WP +BAO are combined with the results of several surveys which detect the BAO signal in the power spectrum:

6dF Galaxy survey: Constraints on $r_{\rm s}(z_{\rm d})/D_{\rm V}(z_{\rm eff}) = 0.336 \pm 0.015$ (4.5% precision) and $D_{\rm V}(z_{\rm eff}) = (456 \pm 27)\,{\rm Mpc}$ where $z_{\rm eff} = 0.106$ [56; 197].

SDSS DR7: Reanalysed Sloan Digital Sky Survey (SDSS) data release (DR) 7 constraints by [275], with $D_{\rm V}(z_{\rm eff})/r_{\rm s}(z_{\rm d}) = 8.88 \pm 0.17$ at $z_{\rm eff} = 0.35$.

SDSS BOSS DR9: The Baryon Oscillation Spectroscopic Survey (BOSS) maps the spatial distribution of luminous red galaxies and quasars to detect the characteristic BAO scale. The results constrain $D_{\rm V}(z_{\rm eff})/r_{\rm s}(z_{\rm d}) = 13.67 \pm 0.22$ at $z_{\rm eff} = 0.57$ [27].

*Planck*2015+Pol+BAO also contains the results from the 6dF Galaxy Survey [56] and the SDSS DR7 Main Galaxy Sample [275] but the updated final SDSS-III BOSS DR12 CMASS and LOWZ [143] results are used where $H(z_{\rm LOWZ})r_{\rm s}(z_{\rm d}) = (11.63\pm0.69)\times10^3{\rm km\,s^{-1}}$ and $D_{\rm A}(z_{\rm LOWZ})/r_{\rm s}(z_{\rm d}) = 6.67\pm0.15$ with $r_{\rm HD_A} = 0.35$ for the LOWZ sample ($z_{\rm LOWZ} = 0.32$) and $H(z_{\rm CMASS})r_{\rm s}(z_{\rm d}) = (14.67 \pm 0.42) \times 10^3{\rm km\ s^{-1}}$ and $D_{\rm A}(z_{\rm CMASS})/r_{\rm s}(z_{\rm d}) = 9.47 \pm 0.12$ with $r_{\rm HD_A} = 0.52$ for the CMASS sample ($z_{\rm CMASS} = 0.57$) [142]. Here the angular diameter distance $D_{\rm A}(z_{\rm eff})$ is constrained instead of the volume-averaged distance.

The *Planck*2013 data is used when considering extensions to ΛCDM, whilst *Planck*2015 data is used for discussion on the quantification of discordance. It will be seen in the subsequent discussion there are subtle quantitative differences between the conclusions that are drawn by choosing a particular CMB data set, but the qualitative results are the same.

2.2 Large scale structure

LSS can be measured via a number of different probes. Four independent measurements of LSS are considered here, each of which can be consistently combined to form a total constraint denoted *All LSS*. Since LSS cannot constrain the optical depth to reionisation it is fixed to the central value from *Planck*2015+Pol+BAO of $\tau_{\rm R} = 0.078$, unless the results are compared with *Planck*2013+WP+BAO or WMAP+highL+BAO in which case $\tau_{\rm R} = 0.09$. An important point to note is the prior ranges set on the parameter ranges for each of the data sets. When using either *Planck*2013+WP+BAO or WMAP+highL +BAO uniform priors of $\Omega_{\rm b}h^2 = [0.005, 0.1]$, $\Omega_{\rm c}h^2 = [0.01, 0.99]$ and $\ln(10^{10}A_{\rm s}) = [2.7, 4]$ as well as well determined Gaussian priors of $100\Theta_{\rm MC} = 1.04131\pm0.00063$ and $n_{\rm s} = 0.9603 \pm 0.0073$ [9] are applied to the LSS data sets. Alternatively, when using *Planck*2015+Pol+BAO the same wide priors as used in the thorough analysis of weak lensing using the Canada France Hawaii Lensing Survey (CFHTLenS) [199] are adopted, $\Omega_{\rm b}h^2 = [0.013, 0.033]$, $\Omega_{\rm c}h^2 = [0.01, 0.99]$, $\Theta_{\rm MC} = [0.5, 10]$, $n_{\rm s} = [0.7, 1.3]$ and $\ln(10^{10}A_{\rm s}) = [2.3, 5]$. The application of different priors leads to quite different outcomes in terms of quantification of

the level of agreement between the posterior distributions found using either the CMB or LSS. This will be discussed in Part III–2.3.

2.2.1 Weak gravitational lensing

In a similar way to using the temperature to calculate the CMB power spectrum, the lensing power spectrum involves extracting the lensing potential across the sky [218; 219; 236]. For weak lensing the Newtonian gravitational potential Φ can be defined by using the Newtonian gauge in cosmological perturbation theory, where $\delta g_{00} = -2a^2\Psi$ and $\delta g_{ij} = 2a^2\Phi\delta_{ij}$ in equation (I–2.11) [109, Chapter 4; 243]. There is no anisotropic stress when $\Psi = \Phi$ [109, Chapter 4; 243]. Through equation (I–2.13) the lensing potential can be related to the matter in a ΛCDM universe, which in Fourier space is ([185; 218; 219; 236])

$$k^2\Phi(\mathbf{k}) = -\frac{3}{2}\Omega_{\rm m}H_0^2a^{-1}\delta(\mathbf{k}) \tag{III–2.4}$$

where $\Omega_{\rm m}$ is the matter density parameter, H_0 is the Hubble constant, $\mathbf{k}$ is the wavevector with magnitude k in direction $\hat{n}$ and a in the scale factor where $a = 1$ today. $\delta(\mathbf{k})$ is the Fourier transform of the density contrast at wavevector $\mathbf{k}$. Integrating this along the line-of-sight gives the lensing potential, ψ, at an angle, $(\theta, \varphi) = \hat{n}$, on the sky [219]. An amplification matrix $\mathbf{A}$, describing the mapping of the coordinates of a lensed image to its unlensed source can be written in terms of this lensing potential ([218])

$$A_{ij} = \delta_{ij} - \partial_i\partial_j\psi. \tag{III–2.5}$$

Further, this amplification matrix can be decomposed into

$$\kappa = \frac{1}{2}\nabla^2\psi, \tag{III–2.6}$$

where ∇^2 is the Laplacian $(\partial_\theta\partial_\theta + \partial_\varphi\partial_\varphi)$ and

$$\gamma_1 = \frac{1}{2}(\partial_\theta\partial_\theta - \partial_\vartheta\partial_\vartheta)\psi, \tag{III–2.7}$$

$$\gamma_2 = \partial_\theta\partial_\varphi\psi, \tag{III–2.8}$$

where it is usual to combine $\gamma_1 + i\gamma_2 = \gamma$ [218]. κ and γ are the convergence and shear which provide isotropic and anisotropic magnification of lensed images [218; 219]. As well as anisotropic magnification, γ describes the shape distortion of the images of sources [218; 219]. When the anisotropic stress is zero (as when $\Psi = \Phi$) then the only distortion of the shape of images must come from the gravitational tidal field and so the comic shear γ, totally encapsulates this [219].

As with the temperature fluctuations of the CMB, the lensing potential can be decomposed into functions on a sphere ([185; 219; 236; 357])

$$\psi(\hat{n}) = \sum_{\ell m} \psi_{\ell m} Y_\ell^m(\hat{n}), \tag{III–2.9}$$

where $\ell = 0 \to \infty$ and $m = -\ell \to \ell$ and

$$\psi_{\ell m} = \int d\Omega_{\hat{n}} \psi(\hat{n}) Y_\ell^{m*}(\hat{n}). \tag{III–2.10}$$

The lensing power spectrum is then given, for redshift bins $i, j = 1, 2, \cdots$, by [219; 357]

$$\langle \psi_{\ell m,i} \psi^*_{\ell' m',j} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_{ij}^{\psi}(\ell). \tag{III–2.11}$$

$C_{ij}^{\psi}(\ell)$ is proportional to the matter power spectrum $P_{\rm m}$ by integration along the line-of-sight, including a factor of $3\Omega_{\rm m} H_0^2/2$ [219].

Galaxy lensing

Galaxy surveys measure the statistical shapes of galaxies. These can be related to the lensing potential through the shear and convergence power spectra [169]. The shear power spectrum can be related to the lensing power spectrum by

$$C_{ij}^{\gamma}(\ell) = \frac{1}{4} \left(\frac{(\ell+2)!}{(\ell-2)!} \right)^2 C_{ij}^{\psi}(\ell), \tag{III–2.12}$$

where the details of the calculation are not important here, but arise due to the relation between the lensing potential to the shear via the Jacobi matrix [219]. It is most common to work with the shear correlation functions ξ_+ and ξ_- since they can be measured directly from the galaxy shape catalogues [204]. This means that the shape of galaxies viewed in galaxy surveys are directly related to the cosmic shear. The correlation functions are given by

$$\xi_+(\hat{n} \cdot \hat{n}') = \frac{1}{4\pi} \sum_{\ell=2}^{\infty} (2\ell+1) C^{\gamma} P_\ell(\hat{n} \cdot \hat{n}') \tag{III–2.13}$$

where $P_\ell(\hat{n} \cdot \hat{n}')$ are the Legendre polynomials [219]. The flat sky power spectrum $P^{\gamma}(\ell)$, analogue of equation (III–2.12) can be used when the correlations are over scales where the curvature of the sky is less important, which gives the correlation functions

$$\xi_+ = \frac{1}{2\pi} \int d\ell \ell J_0(\ell\ \hat{n} \cdot \hat{n}') P^{\gamma}(\ell), \tag{III–2.14}$$

and

$$\xi_- = \frac{1}{2\pi} \int d\ell \ell J_4(\ell\ \hat{n} \cdot \hat{n}') P^{\gamma}(\ell), \tag{III–2.15}$$

where J_0 and J_4 are the Bessel functions of the first kind at order 0 and 4 [218; 219]. Measurements of the shear correlation functions therefore reveal information on $\Omega_{\rm m}$ directly through the relation to C^{ψ}.

The convergence power spectrum is calculated in a similar way to the cosmic shear power spectrum and is related to the lensing power spectrum by ([219])

$$C_{ij}^{\kappa} = \frac{1}{4}\left(\frac{(\ell+1)!}{(\ell-1)!}\right)^4 C_{ij}^{\psi}. \quad \text{(III–2.16)}$$

From this, it can be seen that the convergence power spectrum is considerably larger than the shear power spectrum on large scales (low ℓ), but comparable at large ℓ [219].

Although the correlation of ellipticity between galaxies reveals information on the shear field, it is contaminated by the intrinsic alignment of galaxies [357]. Galaxies which form near each other will be aligned due to their gravitational pull on each other. The galaxies which are in the same tidal field will also be aligned with each other along the line-of-sight [169]. If the observed shear field is expanded to

$$\gamma^{\rm obs} = \gamma + \gamma^{\rm I}, \quad \text{(III–2.17)}$$

where γ is the true shear field and $\gamma^{\rm I}$ describes the correlated intrinsic alignment of galaxies the correlation functions which need to be considered are

$$\langle \gamma_i^{\rm obs}\gamma_j^{\rm obs}\rangle = \langle\gamma_i\gamma_j\rangle + \langle\gamma_i\gamma_j^{\rm I}\rangle + \langle\gamma_i^{\rm I}\gamma_j\rangle + \langle\gamma_i^{\rm I}\gamma_j^{\rm I}\rangle, \quad \text{(III–2.18)}$$

where only $\langle\gamma_i\gamma_j\rangle = \xi_+$ provides useful constraints on cosmological parameters [170]. The other objects are expected to be small and can be encapsulated by modelling of galactic physics and controlled via uncertainty biases during analysis. Modelling the gravitational lensing signature is difficult since it involves knowing, to a high precision, galaxy dynamics [220].

The galaxy lensing measurements used in this Thesis are:

Lensing2013: When using *Planck*2013+WP+BAO or WMAP+highL+BAO the CFHTLenS tomographic blue galaxy sample is used as the galaxy lensing data. This was shown in [170] to have an intrinsic alignment signal that was consistent with zero. This eliminates the need to marginalise over any additional nuisance parameters. The cosmic shear correlation functions are estimated in six redshift bins, each with an angular range $1.5 < \theta < 35$ arcmin. The power

spectrum on non-linear scales can be corrected using the `Halofit` fitting formulae [337; 350], which has been shown to be accurate enough to use with massive neutrinos [62]. This data set is always combined with the *Planck*2013 lensing data.

CFHTLenS (Strong): This relates directly to the *Min* case in [199] figure 12 which has the strongest assumptions made about astrophysical uncertainties. There are seven angular bins and seven tomographic redshift bins which each have their own uncertainties related to them. These redshift uncertainties are Gaussians about $\Delta z_1 = -0.045 \pm 0.014$, $\Delta z_2 = -0.013 \pm 0.010$, $\Delta z_3 = 0.008 \pm 0.008$, $\Delta z_4 = 0.042 \pm 0.017$ and $\Delta z_5 = 0.042 \pm 0.034$ leaving the last two bins with flat priors of $\Delta z_{6,7} = [-0.1, 0.1]$, keeping all angular scales. There are also tight priors on the amplitude of intrinsic alignments and the intrinsic alignment luminosity and redshift dependence are zero.

CFHTLenS (Weak): As for the CFHTLenS *(Strong)* case, this also comes from [199] where it is denoted *Max*. The astrophysical assumptions are greatly reduced with wide flat priors on intrinsic alignment measurements and $\Delta z = [-0.1, 0.1]$ for each of the seven tomographic bins, while non-linear scales are cut in the matter power spectrum. The cut to the non-linear scales is the main cause for measurements from CFHTLenS *(Weak)* being much less constraining than CFHTLenS *(Strong)*.

DES Science Verification: The results from the Dark Energy Survey (DES) follow the prescription in [1] where the range of angular scales included is less than in either of the CFHTLenS analyses for each of its three redshift bins. Here uncertainties in the redshift bins are not taken into account and intrinsic alignments are set to zero. As such the constraints are not as tight as the CFHTLenS *(Strong)* but provide a stronger constraint than CFHTLenS *(Weak)*.

Kilo-Degree Survey: During the preparation of this Thesis the Kilo-Degree Survey (KIDS) [226] has produced results which are similar in many ways to those produced by CFHTLenS. Given this, a value for the discordance has not been quoted for this data in the next section, presuming it to be close to that for CFHTLenS.

CMB lensing

Measuring the gravitational lensing of CMB photons can also provide information about cosmological shear correlations related to the matter power spectrum,

hence revealing information about $\Omega_{\rm m}$ and σ_8 [236]. As lensing maps the temperature $\Theta(\hat{n}) \rightarrow \Theta(\hat{n} + \nabla\psi)$ throughout space, the effects of lensing on the CMB power spectrum can be calculated [236]. The correlation of the lensed temperature is given by [185]

$$\begin{aligned} C_\ell^{\rm lensed} &= \left[1 - \frac{\ell^2}{4\pi}\int d\ell' \ell'^3 C_{\ell'}^{\psi}\right] C_\ell \\ &+ \int \frac{d\hat{n}'}{(2\pi)^2}\left[(\hat{n}' - \hat{n})\cdot\hat{n}'\right] C_{|\hat{n}-\hat{n}'|} C_{\hat{n}}^{\psi} \end{aligned} \tag{III–2.19}$$

where C_ℓ is the CMB temperature power spectrum from equation (I–2.22) and C_ℓ is the lensing power spectrum from equation (III–2.11). This represents a convolution of the unlensed temperature power spectrum with the lensing power spectrum to flatten peaks and shift power from large scales to smaller scales in the temperature power spectrum [236]. Equation (III–2.19) shows that the measurements of the temperature anisotropies of the CMB are closely linked to the lensing potential the photons travel through to be observed. By disentangling the lensed spectrum from the unlensed temperature therefore allows information to be learned about the matter power spectrum and the density of matter. The CMB lensing data used is:

Planck2013 lensing: Lensing2013 constraints of cosmic shear are always combined with the gravitational lensing of the CMB using reconstructions from *Planck*2013 [10] and SPT [129]. The combination of *Planck*2013 lensing and Lensing2013 will simply denoted Lensing for convenience.

Planck2015 lensing: The updated *Planck*2015 lensing contains measurements of the lensing power spectrum between $40 < \ell < 400$ as in [15] where other scales are cut due to spurious features.

2.2.2 Redshift-space distortions

Non-linear effects from the peculiar velocities of galaxies within galaxy clusters can be measured by surveys in redshift-space. In particular, pancake-like structures can form in redshift-space due to the velocities of galaxies falling towards the centre of galaxy clusters [191]. Since RSD are the result of perturbations in the velocity field of matter within galaxies the non-linear evolution, galaxy bias and distortion need to be calculated to, at least, second order in perturbation theory [58]. The anisotropic galaxy power spectrum is given by

$$\begin{aligned} P_{\rm g}(k,\mu) &= e^{-(f\mu\sigma_v)^2}\left[P_{{\rm g},\delta\delta}(k) + 2f\mu^2 P_{{\rm g},\delta\theta}(k)\right. \\ &\left. + f^2\mu^2 P_{\theta\theta}(x) + b_1^3 A(k,\mu\beta) + b_1^4 B(k,\mu,\beta)\right]. \end{aligned} \tag{III–2.20}$$

$P_{\mathrm{g},\delta\delta}(k)$, $P_{\theta\theta}(k)$ and $P_{\mathrm{g},\delta\theta}(k)$ are the power spectrum of the density fields, velocity divergence and their cross correlation and are related to the Kaiser terms [202; 351] which enhances the view of the overdensity in redshift space. $A(k,\mu,\beta)$ and $B(k,\mu,\beta)$ are high order correlations between these Kaiser terms and velocity fields [351]. f is the growth factor, k the wavenumber and μ is the cosine of the angle between the line-of-sight and the wavevector. σ_v is the velocity dispersion and $\beta = f/b_1$. The biases are free parameters in the model where b_1 is the renormalised linear bias. The density-density and density-velocity power spectrum also contains higher order biases from the second and third order non-local spectrum and the second order local spectrum, $b_{\mathrm{s}2}$, $b_{3\mathrm{nl}}$ and b_2. The non-local biases can be related to any initial linear order bias such that $b_{\mathrm{s}2} = -4(b_1 - 1)/7$ and $b_{3\mathrm{nl}} = 32(b_1 - 1)/315$ [38; 80; 313]. The anisotropic galaxy power spectrum can then be used to calculate the multipole power spectrum

$$P_\ell(k) \quad = \quad \frac{2\ell+1}{2\alpha_\perp^2\alpha_\parallel}\int_{-1}^{1} d\hat{n}' P_{\mathrm{g}}(k',\hat{n}')P_\ell(\hat{n}\cdot\hat{n}'), \qquad \text{(III–2.21)}$$

where k' and $\hat{n}'$ are the true wavenumber and true angle along the line-of-sight [41], $P_\ell(\hat{n}\cdot\hat{n}')$ are the Legendre polynomials and $\alpha_\perp$, $\alpha_\parallel$ are the scaling factors introduced in equations (III–2.2) and (III–2.3). The joint growth of structure and amplitude of density perturbations of dark matter $f\sigma_8$, can also be constrained using the relative amplitudes of the RSD monopole and quadrupole [58].

Measuring the deviation of observations from a fiducial cosmology allows the RSD to be quantified by the Alcock-Paczynski factor F_{AP}, which is related directly to the fiducial Hubble parameter $H^{\mathrm{fid}}(z)$, and the angular diameter distance $D_{\mathrm{A}}^{\mathrm{fid}}(z)$ for a given cosmology

$$F_{\mathrm{AP}}(z_{\mathrm{eff}}) \quad = \quad \frac{\alpha_\perp}{\alpha_\parallel}(1+z)D_{\mathrm{A}}^{\mathrm{fid}}(z)H^{\mathrm{fid}}(z)\,. \qquad \text{(III–2.22)}$$

Using $F_{\mathrm{AP}}(z_{\mathrm{eff}})$ and the measure of the BAO signal from equation (III–2.1) the degeneracy between $D_{\mathrm{A}}(z)$ and $H(z)$ can be broken [58]. The data sets used for parameter constraints are:

SDSS-III BOSS DR11 RSD: The RSD parameters and their covariance estimated with $k_{\mathrm{max}} = 0.20h\,\mathrm{Mpc}^{-1}$ are ([58])

$$\begin{pmatrix} D_{\mathrm{V}}(z_{\mathrm{eff}})/r_{\mathrm{s}}(z_{\mathrm{d}}) \\ F_{\mathrm{AP}}(z_{\mathrm{eff}}) \\ f(z_{\mathrm{eff}})\sigma_8(z_{\mathrm{eff}}) \end{pmatrix} \quad = \quad \begin{pmatrix} 13.88 \\ 0.683 \\ 0.422 \end{pmatrix}, \qquad \text{(III–2.23)}$$

$$C^{-1}_{k_{\rm max}=0.20} = \begin{pmatrix} 31.032 & 77.773 & -16.796 \\ & 2687.7 & -1475.9 \\ & & 1323.0 \end{pmatrix}, \tag{III–2.24}$$

with $z_{\rm eff} = 0.57$.

SDSS-III BOSS DR12 RSD: Measurements of the clustering of galaxies along the line-of-sight at effective redshifts of $z_{\rm LOWZ} = 0.32$ and $z_{\rm CMASS} = 0.57$ can constrain $f\sigma_8$ and the combination of the Hubble parameter and the comoving sound horizon at the baryon drag epoch $H(z)r_{\rm s}(z_{\rm d})$, and the ratio of the angular diameter distance to the sound horizon $D_{\rm A}(z)/r_{\rm s}(z_{\rm d})$ ([143])

$$\begin{pmatrix} f(z_{\rm LOWZ})\sigma_8(z_{\rm LOWZ}) \\ H(z_{\rm LOWZ})r_{\rm s}(z_{\rm d}) \\ D_{\rm A}(z_{\rm LOWZ})/r_{\rm s}(z_{\rm d}) \end{pmatrix} = \begin{pmatrix} 0.392 \\ 11.48 \times 10^3 {\rm km\ s^{-1}} \\ 6.38 \end{pmatrix}, \tag{III–2.25}$$

$$\begin{pmatrix} f(z_{\rm CMASS})\sigma_8(z_{\rm CMASS}) \\ H(z_{\rm CMASS})r_{\rm s}(z_{\rm d}) \\ D_{\rm A}(z_{\rm CMASS})/r_{\rm s}(z_{\rm d}) \end{pmatrix} = \begin{pmatrix} 0.445 \\ 13.99 \times 10^3 {\rm km\ s^{-1}} \\ 9.43 \end{pmatrix}. \tag{III–2.26}$$

Here, the covariance matrix for these parameters from the Quick-Particle-Mesh (QPM) mocks are used

$$C^{-1}_{\rm LOWZ} = \begin{pmatrix} 669.33 & -34.364 & -96.193 \\ & 6.2323 & -1.9444 \\ & & 54.992 \end{pmatrix}, \tag{III–2.27}$$

$$C^{-1}_{\rm CMASS} = \begin{pmatrix} 1736.7 & -67.932 & -185.18 \\ & 10.036 & -3.5814 \\ & & 76.115 \end{pmatrix}. \tag{III–2.28}$$

These are calculated from a minimum scale of $k_{\rm max} = 0.24h{\rm Mpc}^{-1}$ to a largest scale of $k_{\rm min} = 0.02h{\rm Mpc}^{-1}$ for the monopole, and $k_{\rm min} = 0.04h{\rm Mpc}^{-1}$ for the quadrupole.

In figure III–2.1 the theoretical RSD multipole power spectra are computed according to the same procedure as in [58]. First, corrections to the linear `CAMB` [237] power spectrum are applied at two-loop order using `RegPT` [352]. The anisotropic galaxy power spectrum can then be modelled using the prescription in [351], which includes corrections due to the coupling between the density and velocity components. Bias corrections are then applied to the density

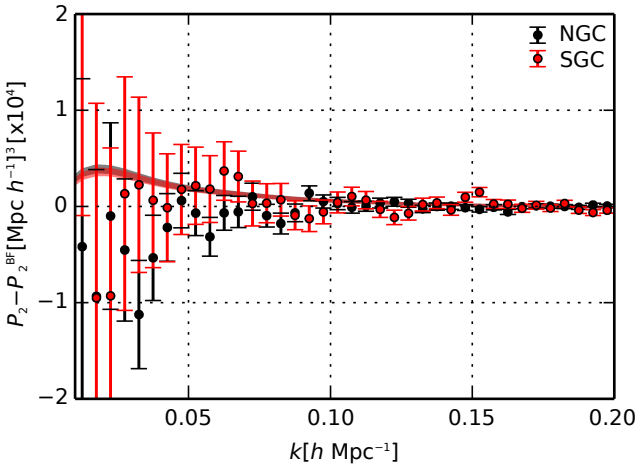


Figure III–2.1: The quadrupole component of the redshift space power spectrum for the range of allowed *Planck*2013 cosmologies (indicated by the narrow red and black bands), relative to the best-fit BOSS spectrum, for the NGC and SGC.

field according to [258], and finally, window functions for the north galactic cap (NGC) and south galactic cap (SGC) added. The plot in figure III–2.1 shows the quadrupole component of the anisotropic power spectrum for *Planck*2013 ΛCDM cosmologies, relative to the best-fit BOSS spectra (where the growth rate is treated as a free parameter) for the NGC and SGC. For the purposes of the plot, the bias parameters are fixed (which are nuisance parameters in the full fit) to their BOSS best-fit values, but included in the range of *Planck*2013 cosmologies allowed by an MCMC analysis. The excess power (on large scales) is apparent for the *Planck*2013 cosmologies, visually showing the preference of the BOSS data for a lower growth rate. In practice calculating the non-linear corrections are computationally expensive, so the SDSS-III BOSS DR11 RSD data are used for MCMC fitting.

The intermediate data products $\{D_{\rm V}(z_{\rm eff})/r_{\rm s}(z_{\rm d}), F_{\rm AP}(z_{\rm eff}), f(z_{\rm eff})\sigma_8(z_{\rm eff})\}$ or $\{f(z_{\rm eff})\sigma_8(z_{\rm eff}), H(z_{\rm eff})r_{\rm s}(z_{\rm d}), D_A(z_{\rm eff})/r_{\rm s}(z_{\rm d})\}$ are used in the MCMC analysis. The RSD results have also been shown to be accurate enough to use with models with non-zero neutrino mass [57]. It should be noted that BAO are never combined with RSD results since they are correlated.

2.2.3 Sunyaev–Zel'dovich galaxy cluster counts

Inverse Compton scattering of CMB photons by high energy electrons in intracluster media can be used to measure the number of galaxy clusters as a function of redshift, from which the growth of structure and various geometrical factors

can be constrained [17]. In particular, the number of galaxy clusters in a solid angle and redshift interval is given by

$$\frac{dN}{d\Omega dz} = \frac{dV}{d\Omega dz}\int dM \frac{dn}{dM} \tag{III–2.29}$$

where the term before the integral describes the comoving volume which is related directly, via general relativity, to the energy density in the universe [77]. The comoving density is

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}}\frac{\varrho_{\rm m}}{M}\frac{\delta_{\rm col}}{\sigma^2}\frac{d\sigma}{dM}\exp\left[-\frac{\delta_{\rm col}^2}{2\sigma_M^2 D^2}\right]. \tag{III–2.30}$$

M is the mass of the cluster, $D(z)$ is the growth of density perturbations as a function of redshift and $\varrho_{\rm m}$ is the density of matter today. σ is the RMS density fluctuation at a given mass scale and redshift and σ_M is the same calculated at $z = 0$. $\delta_{\rm col}$ is the threshold at which an overdensity will collapse and is $\delta_{\rm col} \approx 1.68$ [294]. This result has been updated slightly by [331], but the exact details are not needed here. Importantly, by knowing the number of galaxy clusters in any direction, a direct understanding the amount of matter in the universe can be garnered [154].

The actual detection of the galaxy cluster via the SZ effect is evident from the increase in energy of CMB photons when they pass through high energy electrons. For simplicity, considering one photon and one high energy electron, the energy E, of the incoming photon gets shifted $E' \to E(1+\beta\cos(\theta))$ where β is the ratio of the velocity of the electron to the speed of light and θ is the angle between the electron's velocity and the photon's path [276]. CMB photons passing through the intra-cluster medium will therefore deflect off the high energy photons, changing their paths and increasing in energy. This can be detected as a shifting of the photon temperature in observations of the CMB [12].

The relationship between the observable SZ flux, Y, and the mass of the cluster, M, must be determined empirically using either observations or simulations. A simple assumption for the thermal state of a cluster is to assume hydrostatic equilibrium [17]. Any deviation from the $Y - M$ relation derived from this assumption is quantified using a hydrostatic mass bias $1 - b$. This factor can be constrained using follow-up observations of X-ray detected samples using weak lensing or directly from the lensing effect of clusters on the CMB, measured from the *Planck* data.

Planck2013 lensing: *Planck* has detected clusters via the SZ effect. Using a sample of 189 clusters, cosmological constraints can be deduced in the

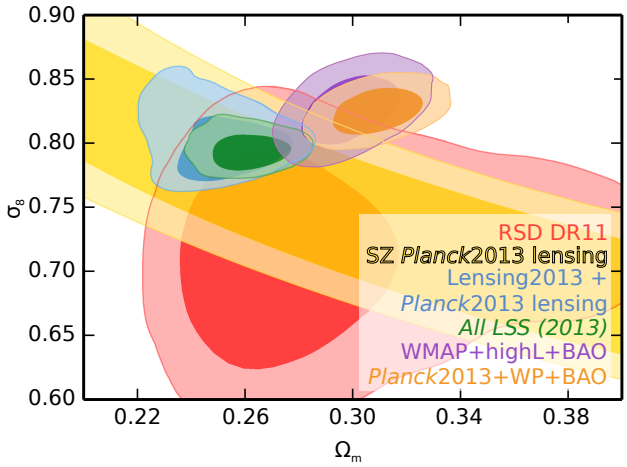


Figure III–2.2: 1σ and 2σ constraint contours in the $\Omega_m - \sigma_8$ plane within the ΛCDM model for a range of data. RSD DR11, SZ Planck2013, Lensing2013+Planck2013 lensing are plotted in red, yellow and blue respectively. Combining each of the independent LSS constraints in *All LSS (2013)* forms the green contours. WMAP+highL+BAO and *Planck*2013+WP+BAO CMB constraints are in purple and orange.

$\sigma_8 - \Omega_m$ plane [11]. The constraints can then be implemented by imposing priors on $\sigma_8(\Omega_m/0.27)^{0.3} = 0.764 \pm 0.025$, where $1-b$ is allowed to vary in the range $[0.7, 1.0]$. This is compatible with other determinations using cluster counts selected using the SZ effect [162; 319] and in other wavebands [308; 368].

Planck2015 lensing: The lensing effect of 439 clusters on the CMB can be used to infer $1/(1-b) = 0.99 \pm 0.19$ [15; 17]. This greatly improves the constraint from *Planck*2013 lensing.

Weighing the Giants: There are 51 galaxy clusters in the sample studied by the Weighing the Giants (WtG) project, 22 of which overlap with the *Planck*2015 galaxy clusters, for which lensing data exists [246]. The mass bias determined by WtG is lower than for *Planck*2015, at $1 - b = 0.688 \pm 0.092$, and as such galaxy cluster dynamics suggest that these objects deviate significantly from hydrostatic equilibrium.

When using *Planck*2015 lensing or WtG to calculate the bias for SZ galaxy cluster counts, the BAO combinations mentioned in Part III–2.1 are used, unless the SZ constraints are combined with RSD.

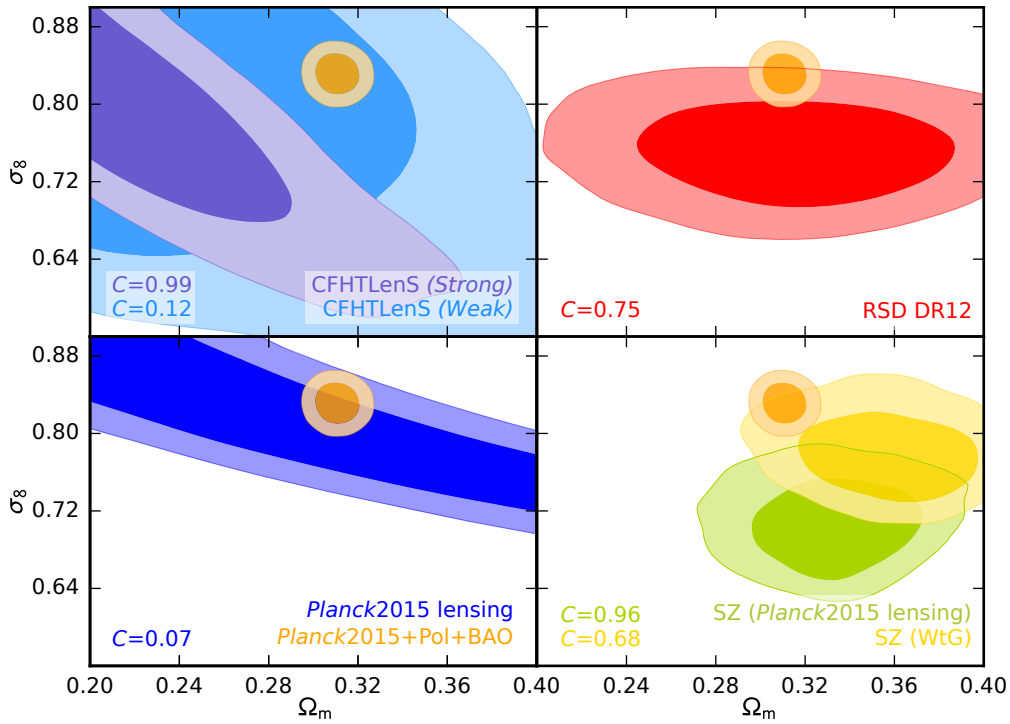


Figure III–2.3: 1σ and 2σ constraint contours in the $\Omega_m - \sigma_8$ plane within the ΛCDM model for a range of data. In each subplot the orange contours show the constraints from *Planck*2015+Pol+BAO. The top left subplot shows the constraints from weak lensing with the CFHTLenS *(Strong)* and CFHTLenS *(Weak)* results plotted in purple and and light blue respectively. The bottom left and top right subplots show the constraint from CMB lensing in dark blue and from BOSS DR12 RSD in red. The bottom right subplot contains the constraint from SZ galaxy cluster counts with mass biases from CMB lensing in lime green and WtG in yellow.

2.3 Parameter constraints

To get a description of the posterior distribution on the five relevant ΛCDM parameters $\theta = \{\Omega_b h^2, \Omega_c h^2, 100\Theta_{MC}, \ln(10^{10} A_s), n_s\}$ MCMC analysis can be performed using either LSS data or the CMB (as described in Part I–3.4). This gives samples which can then be used to learn about statistics such as expected parameter values, correlations between parameters and more. The samples can also be used directly to assess differences in the posterior distributions of cosmological parameters given different data sets. Following [9], the sum of the masses of the active neutrinos are fixed to $\sum m_\nu = 0.06\text{eV}$ within the ΛCDM model in order to satisfy the results from oscillation experiments.

Plotted in figures III–2.2 and III–2.3 are the $\Omega_m - \sigma_8$ 1σ and 2σ constraint

contours projected from constraints on the ΛCDM parameters for a range of LSS distributions as well as the three different CMB constraints. Figure III–2.2 shows the comparison of the 2013 data sets where it can be seen that each of the independent LSS constraint contours lie slightly outside of the CMB contours, but when all of the LSS constraints are combined into *All LSS (2013)* the green 2σ contours only just touch the *Planck*2013+WP+BAO orange contours. More explicitly, in figure III–2.3 each of the LSS constraint contours are plotted separately (comparing CFHTLenS *(Strong)* and *(Weak)* or SZ (*Planck*2015 lensing) and SZ (WtG) in the top left and bottom right subplots) with *Planck*2015+Pol +BAO in orange. In the bottom left of each subplot of figure III–2.3 the quantification of tension using the difference vector method (*3*) from Part III–1 is plotted. This quantification of tension is for the five dimensional parameter space containing each of the relevant ΛCDM parameters, and for not the discordance in the $\Omega_{\rm m} - \sigma_8$ plane. This means that the C values quoted may not be reflected by the visual examples in the plots. Table III–2.1 shows the plotted C values and their interpretation when mapped on to the intervals of a 1D Gaussian (the DES scientific verification results are also included in the table).

The discordances between the posterior distributions using *Planck*2015+Pol+ BAO and each of the LSS data sets are not particularly significant, with the possible exceptions of CFHTLenS *(Strong)* and SZ (*Planck*2015 lensing) - both of which are barely significant. Since each of the LSS probes are independent measurements they can be combined to provide an *All LSS* constraint. If each of the mildly discrepant LSS constraints lie in the same region of parameter space, then their combination can become more significant than each separately. Three combinations considered here are:

All LSS (2013): The combined Lensing2013 and *Planck*2013 lensing results, DR11 RSD measure and SZ (*Planck*2013 lensing) galaxy cluster counts are combined. Gaussian priors of $100\Theta_{\rm MC} = 1.04131 \pm 0.00063$ and $n_{\rm s} = 0.9603 \pm 0.0073$ are also applied during the MCMC analysis. Comparison of this constraint (green) to *Planck*2013+WP+BAO (orange) and WMAP+highL+BAO (purple) are shown in figure III–2.2. Using the difference vector method of quantification of the level of agreement between the posterior distributions of *All LSS (2013)* and *Planck*2013+WP+BAO and mapping the result directly to an interval on a 1D Gaussian there is a tension which exceeds 5σ. This is not reflected in the $\Omega_{\rm m} - \sigma_8$ contours of figure III–2.2 because these derived parameters are not highly correlated in the ΛCDM parameter directions where the tension arises. More on this result will be discussed at the end of this section.

Comparison data set	Difference vector	Interpretation
CFHTLenS *(Strong)*	0.99	2.7σ
CFHTLenS *(Weak)*	0.12	0.15σ
DES scientific verification	0.62	0.90σ
*Planck*2015 lensing	0.07	0.08σ
RSD DR12	0.75	1.2σ
SZ (*Planck*2015 lensing)	0.96	2.0σ
SZ (WtG)	0.68	0.99σ

Table III–2.1: Amount of discordance between the *Planck*2015+Pol+BAO ΛCDM posterior distributions and each of the LSS data set constraints from figure III–2.3 (also including DES scientific verification results). The *Interpretation* is found by mapping C to an interval on a 1D Gaussian.

All LSS (Weak): By combining CFHTLenS *(Weak)* with *Planck*2015 lensing, RSD DR12 and SZ galaxy cluster counts using the WtG mass bias the least discrepant joint analysis compared to *Planck*2015+Pol+BAO can be found. In figure III–2.4 the green contours in the $\Omega_{\rm m} - \sigma_8$ plane visually overlap substantially with *Planck*2015+Pol+BAO giving a difference vector value $C = 0.55$ (0.76σ). For more information on the correlated tensions in the five relevant ΛCDM parameters the 2D projection for every combination is presented in figure III–2.5.

All LSS (Strong): Combining the CFHTLenS *(Strong)* constraints with RSD DR12, *Planck*2015 lensing and SZ galaxy cluster counts using the mass bias from *Planck*2015 lensing is shown in brown in figure III–2.4. The 2D projections for each of the five relevant ΛCDM parameters can also be found in figure III–2.6. *All LSS (Strong)* is a more discrepant combination of data than *All LSS (Weak)* with $C = 0.99$ (2.55σ). Note that this is less discrepant than the CFHTLenS *(Strong)* discrepancy by itself. This suggests that there are internal tensions between the LSS data sets, as well as with CMB constraints.

The calculated values for each of the statistics introduced in Part III–1 are presented in tables III–2.2 and III–2.3. These indicate that *All LSS (Strong)* is more discrepant than the parameter distributions inferred from *Planck*2015+Pol +BAO whilst using *All LSS (Weak)* appears to be more compatible.

The results of measures *1* and *2* in tables III–2.2 and III–2.3 are small compared to $B = 1$ or $O = 1$ suggesting a large degrees of discordance between

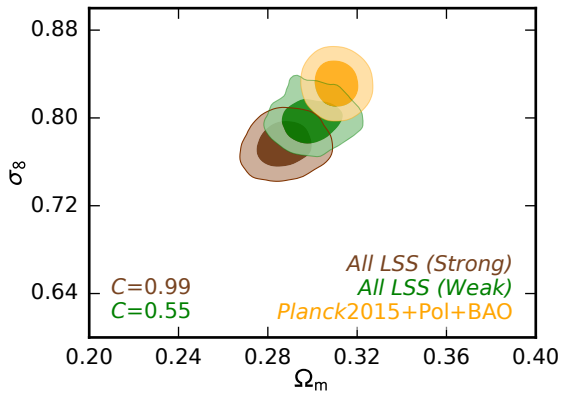


Figure III–2.4: The 1σ and 2σ constraints on the $\Omega_m - \sigma_8$ plane from *Planck*2015+Pol+BAO in orange and from combining each of the LSS data sets, with those in the most tension with the CMB data set in brown and in the least tension in green. The five parameter ΛCDM difference vector with *Planck*2015+Pol+BAO is quoted for both sets of constraints in the bottom-left corner.

Measure	Result			Interpretation
1	B	=	1.81×10^{-2}	Unknown
2	O	=	2.71×10^{-3}	Unknown
3	C	=	$0.55\ (0.76\sigma)$	Low
4	$I_{\rm CMB}$	=	3.81×10^{-1}	Low
	$I_{\rm LSS}$	=	2.30×10^{-3}	
5	$D({\rm CMB}\|\|{\rm LSS})$	=	7.20×10^{-2}	Likely
	$S_{\rm CMB\to LSS}$	=	-4.25×10^{-1}	similar
	$D({\rm LSS}\|\|{\rm CMB})$	=	8.52	
	$S_{\rm LSS\to CMB}$	=	8.03	
6	$\log R$	=	3.29	Low
7	$\log T$	=	2.59	Mild

Table III–2.2: Quantification of the similarity of the probability distributions of the ΛCDM parameters from *Planck*2015+Pol+BAO and *All LSS (Weak)* for each of the measures *1*-*7* from Part III–1. The first column contains the measure used, the second column shows the result and the final column gives a description of degree of discordance.

Measure	Result			Interpretation
1	B	$=$	8.90×10^{-4}	Unknown
2	O	$=$	9.70×10^{-5}	Unknown
3	C	$=$	0.99 (2.55σ)	Moderate
4	I_{CMB}	$=$	2.82×10^{-2}	Moderate
	I_{LSS}	$=$	5.44×10^{-5}	
5	$D(\mathrm{CMB}\|\|\mathrm{LSS})$	$=$	2.85×10^{-3}	Likely
	$S_{\mathrm{CMB}\to\mathrm{LSS}}$	$=$	-5.85	different
	$D(\mathrm{LSS}\|\|\mathrm{CMB})$	$=$	7.84	
	$S_{\mathrm{LSS}\to\mathrm{CMB}}$	$=$	1.99	
6	$\log R$	$=$	-1.36	Significant
7	$\log T$	$=$	7.56	Significant

Table III–2.3: Identical table to Table. III–2.2 using *All LSS (Strong)* to constrain the LSS parameter distributions. The first column contains the measure used, the second column shows the result and the final column gives a description of the degree of discordance.

constraints obtained from LSS and CMB datasets. To illuminate how poor these measures are at quantifying tension, a toy model can be considered to see what the results are equivalent to in terms of shifts of two distributions. If $P_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $P_2 = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}_1 = (0,0,0,0,0)$, $\boldsymbol{\mu}_2 = (0,0,0,0,\theta)$ and $\boldsymbol{\Sigma} = \mathrm{diag}(1,1,1,1,1)$ then $B = 1.81 \times 10^{-2}$ needs $\theta = 4.23$ whilst $B = 8.90 \times 10^{-4}$ needs $\theta = 5.59$. In a similar way $O = 2.71 \times 10^{-3}$ requires $\theta = 3.48$ and $O = 9.70 \times 10^{-5}$ needs $\theta = 4.42$. From these shifts in the five dimensional distributions it appears that *All LSS (Weak)* and *All LSS (Strong)* are both quite distinct from *Planck*2015+Pol+BAO. There is a strong dimensional dependence using these two measures so extremely small values can, and do, appear as large discrepancies. On the basis of this, these measures indicate significant discordance between *All LSS (Weak)* and *Planck*2015+Pol+BAO and severe discordance between *All LSS (Strong)* and *Planck*2015+Pol+BAO. However, since the shift in the means has an equivalent description in terms of broadening of the variance then it is difficult to make any useful statement. Instead consider another toy model where $P_1 = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1)$ and $P_2 = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{\mu} = (0,0,0,0,0)$, $\boldsymbol{\Sigma}_1 = \mathrm{diag}(1,1,1,1,1)$ and $\boldsymbol{\Sigma}_2 = \mathrm{diag}(\sigma^2,\sigma^2,\sigma^2,\sigma^2,\sigma^2)$ then $B = 1.81 \times 10^{-2}$ needs $\sigma \approx 10$ whilst $B = 8.90 \times 10^{-4}$ needs $\sigma \approx 33$. Neither of these P_2 distributions would be considered in tension with P_1, although P_2 would not be informative. In general, the values of B and O are much less than one, which would suggest that there is reasonably significant discordance

between *All LSS (Weak)* or *All LSS (Strong)* and *Planck*2015+Pol+BAO. It is clear why neither the Bhattacharyya distance nor the overlap coefficient measures are used for data comparison in cosmology.

Measure *3* is simple to interpret from both tables III–2.2 and III–2.3. Since the value of C is the fraction of samples within an interval then it maps easily to the number of samples within an interval of a 1D Gaussian distribution. As such C maps directly to a quantification in terms of a number of standard deviations. For *All LSS (Weak)* compared to *Planck*2015+Pol+BAO, $C = 0.55$ is equivalent to 0.76σ, which is interpreted as very little discordance. Comparing *All LSS (Strong)* to *Planck*2015+Pol+BAO provides $C = 0.99$ which (including more significant figures in the calculation) maps to 2.55σ. While this is much greater than for *All LSS (Weak)*, the suggested interpretation of the tension is only moderate. These values reflect the position of the contours in figure III–2.4.

When interpreting measure *4* it is only necessary to consider $\mathrm{Max}[I_{\mathrm{CMB}}, I_{\mathrm{LSS}}]$ to get an indication of the level of agreement. The ratio of the larger value to the smaller value then describes the relative widths of the distributions. In a similar way to measure *3*, the values of I_{CMB} and I_{LSS} relate directly to numbers of samples, although the distributions are cut at the complementary distributions 3σ isocontours, meaning they are discontinuous. While this means they cannot truly be mapped to intervals of a 1D Gaussian, that is still a useful indicator of discordance. For *All LSS (Weak)* 38.1% of the samples drawn from the *Planck*2015+Pol+BAO distribution are within the isocontour which would contain 99.7% of the samples drawn from the *All LSS (Weak)* distribution. This seems like a small fraction, but is actually the equivalent of a discrepancy of 0.88σ when compared to a 1D Gaussian and so should be interpreted as indicating a low level of discordance. Since $I_{\mathrm{CMB}} > I_{\mathrm{LSS}}$ then the constraints on the parameters using *Planck*2015+Pol+BAO are tighter than those from *All LSS (Weak)*. Similarly, I_{CMB} is larger in table III–2.3 showing that the constraints from *Planck*2015+Pol+BAO are tighter than those from *All LSS (Strong)*. $I_{\mathrm{CMB}} = 2.82 \times 10^{-2}$ means that 2.82% of the samples drawn from the *Planck*2015+Pol+BAO distribution are within the isocontour which would contain 99.7% of the samples drawn from the *All LSS (Strong)* distribution. Again, this seems quite low but is the equivalent to 2.2σ and so is again only moderately discordant.

Measure *5*, is a bit more difficult to interpret. In the case of updating both the *All LSS (Weak)* and the *All LSS (Strong)* constraints with the constraints

from *Planck*2015+Pol+BAO there is little relative entropy, but large negative *suprise*. Since the values of the *suprise* are negative this suggests that the distributions are more similar than expected. It should be noted that this does not mean that the distributions are similar, just that there is less of an information gain than expected. Indeed, it is is very difficult to quantify quite how severe the discordance is using this measure; it should rather be used to describe whether one data set is likely to update another data set. The important outcome of the measure *5* results from tables III–2.2 and III–2.3 is the similarity between the results for $D(\mathrm{LSS}||\mathrm{CMB})$ and those for $S_{\mathrm{LSS}\to\mathrm{CMB}}$ for *All LSS (Weak)*. This indicates that the distributions are likely to be similar, but $D(\mathrm{LSS}||\mathrm{CMB})$ being larger than $S_{\mathrm{LSS}\to\mathrm{CMB}}$ shows that it is more probable that the parameter distributions from *Planck*2015+Pol+BAO can be updated with the constraints from *All LSS (Strong)*. This means the distributions are likely to be more distinct.

For measure *6*, table III–2.2 has $\log R = 3.29$ signifying that the joint distribution, with *Planck*2015+Pol+BAO and *All LSS (Weak)* as data sets, is more likely than each of the distributions separately. The similarity is quite significant when using flat priors from the minimum to maximum parameter values obtained in the samples. This happens only when the two distributions are, at worst, mildly discordant. When comparing this to the *All LSS (Strong)* result of $\log R = -1.36$, in table III–2.3, the negative value shows that the joint distribution is less likely than each of the distributions separately, which is true when the distributions are more distinct. It is best to interpret the values of $\log R$ on the Jeffreys scale often used in Bayesian analysis [196], with a result of $\log R = 3.29$ showing *Planck*2015+Pol+BAO is "decisively similar" to *All LSS (Weak)* and $\log R = -1.36$ suggesting *Planck*2015+Pol+BAO is significantly different to *All LSS (Strong)*. These statements are more extreme than the other measures as a result of placing relatively tight priors. Increasing the range of the prior distribution allows less extreme interpretation of the results but with the same quantitative outcome - the *All LSS (Weak)* distribution is more similar to *Planck*2015+Pol+BAO than the *All LSS (Strong)* distribution is.

Finally, measure *7* indicates that the discordance between *All LSS (Weak)* and *Planck*2015+Pol+BAO is mild, but as with measure *6* this statement is prior dependent. Again, the $\log T$ value when using *All LSS (Strong)* shows a much more significant discordance. By changing the priors, the interpretation of this result can change from *All LSS (Weak)* being in almost complete agreement with *Planck*2015+Pol+BAO to there being significant or severe disagreement. The interpretation from *All LSS (Strong)* then follows suit, being always more

discordant than *All LSS (Weak)*.

To summarise the usefulness of each of these methods:

- *1* and *2* cannot give a useful quantification of discordance, although the small values would *suggest* more significant discordance than other methods.
- *3* and *4* can be related to drawn samples from distributions and so mapped to intervals on a 1D Gaussian and tend to give slightly more conservative interpretations of the discordance.
- *5* is useful to find out whether a distribution is likely to usefully update one distribution from a pre-existing data set, but cannot be easily interpreted as a quantification of the difference between the data sets.
- *6* and *7* are prior dependent and so care needs to taken when interpreting the actual values as a indication of the severity of discordance.

All of these measures, for both *All LSS (Weak)* and *All LSS (Strong)*, are not representative of the value of the tension obtained by comparing *All LSS (2013)* and *Planck*2013+WP+BAO. This is true even though the $\Omega_{\rm m} - \sigma_8$ 2σ contour for the *All LSS (Strong)* in figure III–2.4 looks similar to the *All LSS (2013)* contour in III–2.2. This is due to the application of *Planck*2013 priors on $\Theta_{\rm MC}$ and $n_{\rm s}$. These parameters were chosen since they are well measured by the CMB and in particular $\Theta_{\rm MC}$ is known to within 0.05%. Using importance sampling on the *All LSS (Strong)* chains and placing priors of $100\Theta_{\rm MC} = 1.04086 \pm 0.00048$ and $n_{\rm s} = 0.9652 \pm 0.0062$, the resulting tension with *Planck*2015+Pol+BAO is $C = 0.999(95)$ (4.06σ) which is in closer agreement with the *All LSS (2013)* result. There are relatively few samples in the prior regions of $\Theta_{\rm MC}$ and $n_{\rm s}$ when using *All LSS (Strong)* so the probability distribution from the samples is likely not to be representative of the true distribution.

Since these values are restricted to a particular region of their parameter space, the other three ΛCDM parameters ($\Omega_{\rm b}h^2$, $\Omega_{\rm c}h^2$ and $\ln(10^{10}A_{\rm s})$) become constrained to less favourable regions, i.e. distinctly different from *Planck*2013+WP +BAO. Figures III–2.5 and III–2.6 show the projected likelihood contours comparing *Planck*2015+Pol+BAO to *All LSS (Weak)* and *All LSS (Strong)* respectively. Although, not entirely accurate - the application of priors on $\Theta_{\rm MC}$ and $n_{\rm s}$ would restrict the green and the brown contours to the size of the *Planck*2015+Pol+BAO contours in the $\Theta_{\rm MC}$ and $n_{\rm s}$ directions. For figure III–2.5, even though the priors limit the *All LSS (Weak)* parameter distributions in

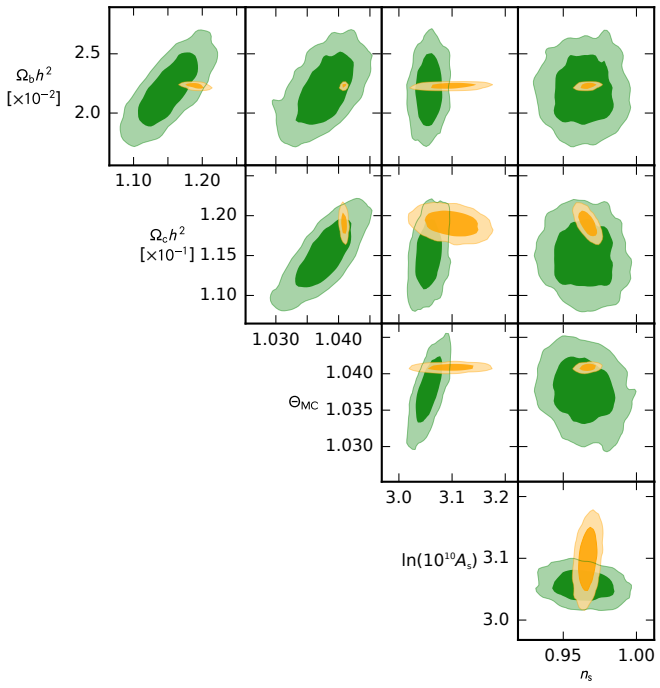


Figure III–2.5: Projected 1σ and 2σ likelihood contours for each of the relevant ΛCDM parameters. The green contours show the *All LSS (Weak)* constraints and the orange contours are the constraints from *Planck*2015+Pol+BAO.

all directions, they do not become significantly more discrepant with *Planck*2015 +Pol+BAO. On the other hand, in figure III–2.6, the priors restrict $\ln(10^{10}A_{\rm s})$ and $\Omega_{\rm c}h^2$ to the upper range of their allowed values. This causes a knock on effect requiring both lower and higher $\Omega_{\rm b}h^2$ values from the correlation with $\Omega_{\rm c}h^2$ and the allowed region from the priors respectively. This "new constraint" lies further from *Planck*2015+Pol+BAO and so the agreement with *All LSS (Strong)* with priors decreases. It should be noted that it is naïve to use the combinations of the 2D contours in figures III–2.5 and III–2.6 to make serious assumptions about shifts in the distributions with the application of priors. The true distributions are five dimensional and can only be projected down to the 2D contours via marginalising out other parameters, therefore losing a lot of information in the process. Releasing these priors to cover a wider range allows more natural values in the remaining parameters to be chosen. The variances of these LSS parameter distributions mean that more samples overlap with samples from the *Planck*2015+Pol+BAO distribution, reducing the tension.

It should be noted that if the belief in the $n_{\rm s}$ and $\Theta_{\rm MC}$ priors is strong, the greater tension result may be favoured. This was the argument made in the 2013 study presented here. The priors themselves come from *Planck*2013 con-

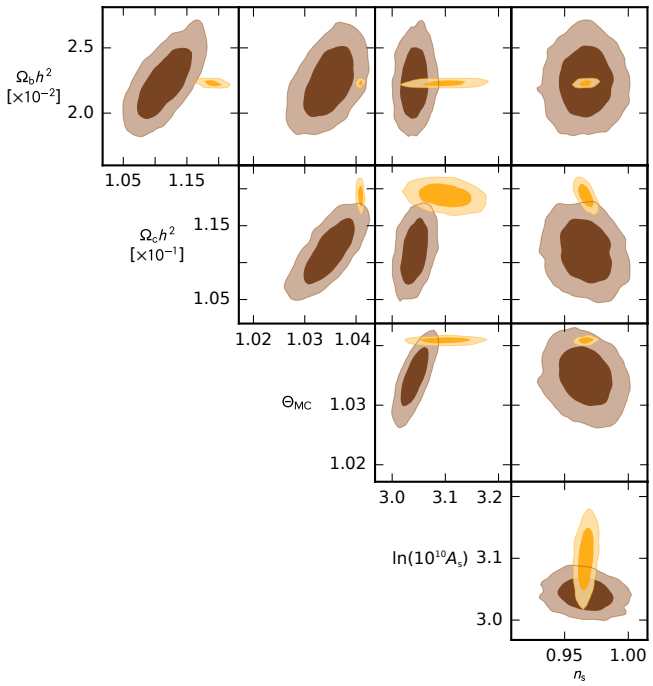


Figure III–2.6: Projected 1σ and 2σ likelihood contours for each of the relevant ΛCDM parameters. The brown contours show the *All LSS (Strong)* constraints and the orange contours are the constraints from *Planck*2015+Pol+BAO.

straints. Even though the priors are well believed, they were are not applied in the 2015 study to decouple the CMB and LSS measurements completely. Neither choice is incorrect, they are both belief dependent.

2.4 ΛCDM extensions

As mentioned in the previous section, if priors are placed on $\Theta_{\rm MC}$ and $n_{\rm s}$ when obtaining samples from the posterior distribution given LSS data then there is significant tension between the distributions from LSS and the CMB. There are various ways to extend ΛCDM to try and alleviate the discordance. A non-exhaustive selection of extensions are laid out below.

2.4.1 Neutrinos

The inclusion of a neutrino component in the cosmological model can reduce the amount of power on small scales for a given large-scale normalisation, $A_{\rm s}$. This is true both in the case of active neutrinos that correspond to the mass eigenstates of the standard three flavours and also for a sterile neutrino, which evades the strong bound on the number of neutrino species from particle physics experiments not coupling to weak interactions. Using *Planck*2015+Pol+BAO

to constrain ΛCDM with either active or sterile neutrinos added indicates no preference for either a mass for active neutrinos, $\sum m_\nu < 0.15$eV or any mass deviation of the number of relativistic degrees of freedom from sterile neutrinos, $m^{\rm eff}_{\rm sterile} < 0.65$eV and $\Delta N_{\rm eff} < 0.342$ [14].

Active neutrinos

The inclusion of active neutrinos is modelled by the addition of a single parameter, $\sum m_\nu$, assuming that this is distributed equally amongst the three species of massive neutrino. This approximates a degenerate hierarchy with $m_1 = m_2 = m_3 = \sum m_\nu/3$, which is true for large $\sum m_\nu$ with respect to the mass differences of the eigenstates, as is the case in the models we will find gives rise to the best fit to the data (see figure III–2.11). Within the currently constrained limits, such models affect structure growth on small scales and the primary anisotropies of the CMB [127; 155; 158; 233].

There is little difference between massive neutrinos (with $\sum m_\nu \lesssim 0.5$ eV) and massless neutrinos in terms of their effect on pre-recombination dynamics - both the background and of perturbations - since they are relativistic at recombination in both cases. The differences that do arise are due to the ratio of the angular diameter distance to last-scattering, $D_{\rm A}(z_*)$, to the sound horizon at last-scattering, $r_{\rm s}(z_*)$, which sets the angular scale of the CMB acoustic peaks. As the mass of the neutrino increases, $D_{\rm A}(z_*)$ decreases, last-scattering appears closer and anisotropies are shifted to larger angular scales [155]. There is a degeneracy in the effect on the CMB primary anisotropies between dark energy density and massive neutrinos in flat space, in addition to a difference in the Hubble constant, but this degeneracy is broken by several effects including the late-time ISW [182; 235; 301]. The CMB primary anisotropies are affected via the back-reaction on the metric perturbations from the stress-energy of neutrino perturbations. The size of the effect on the CMB is $\mathcal{O}[(\sum m_\nu/k_{\rm B}T_\nu)^2]\rho_\nu/\rho_{\rm tot}$ where ρ_ν is the energy density per species of massless neutrino. For neutrino mass scales relevant to this analysis, changes in the CMB should be $\sim 0.1\%$ as found in [155] where they used $\sum m_\nu = 0.37$eV.

Massive neutrinos also reduce structure growth on small scales compared to massless neutrinos. Neutrinos cluster on scales above their free-streaming length - for a non-relativistic transition in matter domination the free-streaming length is $\lambda_{\rm FS}(z) \propto a^{1/2}$ [233], therefore the comoving free-streaming length decreases with time. The growth of structure is reduced since the neutrino Fourier modes inside the comoving horizon at the non-relativistic transition cannot cluster un-

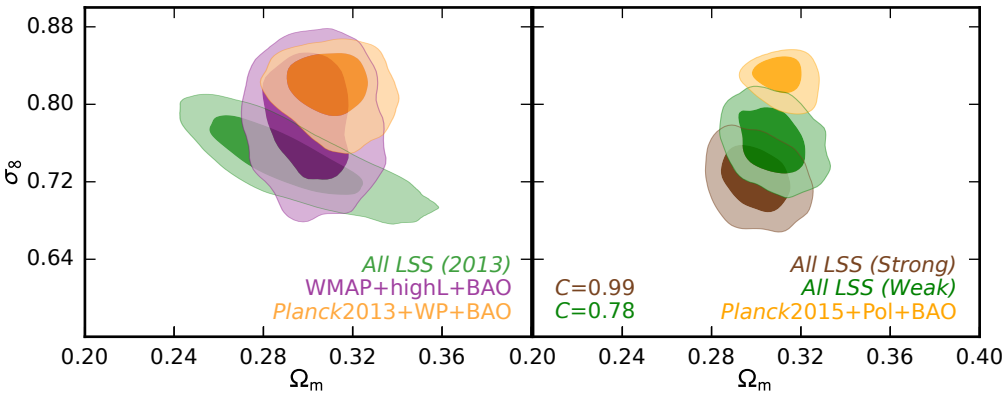


Figure III–2.7: 1σ and 2σ constraints in the $\Omega_m - \sigma_8$ plane for ΛCDM+$\sum m_\nu$ from *Planck*2013+WP+BAO (orange) and WMAP+highL+BAO (purple) on the left and *Planck*2015+Pol+BAO (light orange) on the right and from combining each of the LSS data sets, *All LSS (2013)* (lighter green) on the left and *All LSS (Strong)* (brown) and *All LSS (Weak)* (darker green) on the right. The five parameter ΛCDM difference vector with *Planck*2015+Pol+BAO is quoted for both sets of constraints in the bottom left corner of the right subplot.

til the modes leave the shrinking comoving free-streaming length. There is suppression in the matter power spectrum on smaller scales due to the massive neutrino modes which are currently within the comoving free-streaming length. This is found to have scale-free fractional suppression of $\sim -8f_\nu$ where $f_\nu = \Omega_\nu/\Omega_m$ [186].

The equivalent of figures. III–2.2 and III–2.3 are shown in figure III–2.7 when $\sum m_\nu$ is allowed to vary. There is a significant reduction in the tension between the *All LSS (2013)* constraint (green contours left subplot) and the CMB observations from WMAP+highL+BAO (purple contours). It appears that there is still a tension in the case of *Planck*2013+WP+BAO (orange contours left subplot) although this is weaker than in the case when $\sum m_\nu$ is fixed to 0.06eV. The tension still remains, but now at the level of $\sim 2.5\sigma$. This opposes the right hand subplot of figure III–2.7, particularly in the *All LSS (Strong)* case. Clearly, here, there is very little benefit from adding active neutrinos, evident from the $\Omega_m - \sigma_8$ contours. When including $\sum m_\nu$ in the quantification of discordance analysis, such that the probability distribution is six dimensional $C = 0.999(85)$ (3.79σ) and $C = 0.781$ (1.23σ) for *All LSS (Strong)* and *All LSS (Weak)* respectively. This is due to the distribution of neutrino mass not aligning particularly well when constrained using *All LSS (Strong)* or *(Weak)* compared to *Planck*2015+Pol+BAO and not because an extra degree of freedom has been added. It is interesting that active neutrinos slightly alleviate tension

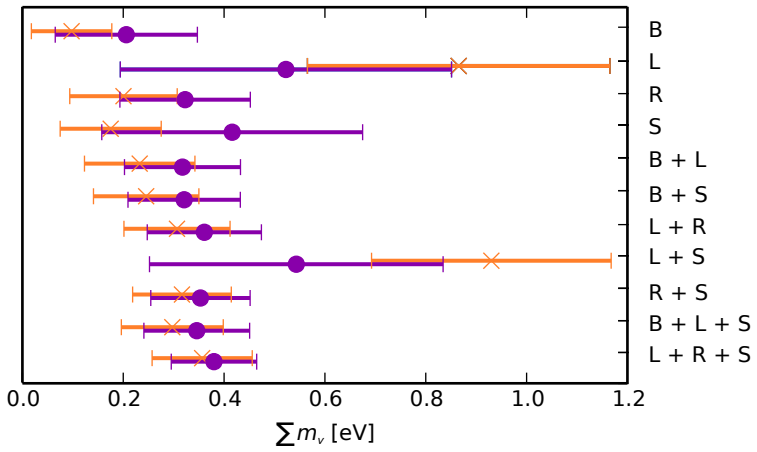


Figure III–2.8: Comparison of the the 1D marginalised value $\sum m_\nu$ and 1σ errors for a wide range of LSS data combinations with CMB data. The CMB data used is *Planck*+WP (orange) or WMAP+highL (purple). The LSS data sets are indicated on the right are labelled B for BAO (where BOSS DR9 is used), L for Lensing2013 + *Planck*2013 lensing, R for RSD DR12 and SZ for *Planck*2013 lensing.

between *All LSS (2013)* and WMAP+highL+BAO posterior distributions (and marginally between *All LSS (2013)* and *Planck*2013+WP+BAO distributions) but worsens the agreement between the posterior distributions using data from *All LSS (Strong)* and *Planck*2015+Pol+BAO. One thing to notice is that priors on $\Theta_{\rm MC}$ and $n_{\rm s}$ already require that constraints from *All LSS (2013)* be similar to *Planck*2013+WP+BAO in two dimensions, and correlation between these parameters and $\sum m_\nu$ forces the distribution of the mass of the neutrinos from both *All LSS (2013)* and *Planck*2013+WP+BAO to be similar. On the other hand, without the priors *All LSS (Strong)* parameters shift away from those constrained by *Planck*2015+Pol+BAO when $\sum m_\nu$ is included, increasing the tension.

It is interesting to consider the distribution of the mass of the active neutrinos for a given range of LSS observations when combined with CMB data. Figure III–2.8 shows the marginalised $\sum m_\nu$ mean values and 1σ errors when combined with *Planck*2013+WP in orange or WMAP+highL in purple. In some cases there is clearly only an upper bound, but as the number of LSS data sets included increases the constraint stabilises to a non-zero value with a significance of around $3-4\sigma$. It is clear that preference for non-zero $\sum m_\nu$ is not dependent on this choice when two or more LSS data sets are included. Moreover, it is clear that there is a preference for non-zero neutrino mass *without* including SZ data. It can

Parameter	*Planck*2013+WP+BAO	*Planck*2013+WP+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02213±0.00025	0.02229±0.00025
$\Omega_{\rm c}h^2$	0.1185±0.0019	0.1154±0.0014
$100\theta_{\rm MC}$	1.04141±0.00057	1.04156±0.00056
$\tau_{\rm R}$	0.092±0.013	0.096±0.014
$n_{\rm S}$	0.9627±0.0061	0.9677±0.0055
$\log(10^{10}A_{\rm s})$	3.090±0.025	3.091±0.027
$\sum m_\nu$[eV]	<0.26	0.357±0.099
H_0	67.57±0.92	66.52±1.15
$\Omega_{\rm m}$	0.311±0.012	0.320±0.015
σ_8	0.818±0.023	0.749±0.019
$z_{\rm re}$	11.21±1.11	11.15±1.20

Table III–2.4: Marginalised parameter table when ΛCDM is extended with active neutrinos. The data combinations are *Planck*2013+WP+BAO in the first column and *Planck*2013+WP+*All LSS (2013)* in the second.

be seen that *Planck*2013+WP with Lensing2013 + *Planck*2013 lensing with or without SZ cluster counts (i.e. excluding BAO data) prefers a higher mass than WMAP+highL, unlike in other cases. This is due to WMAP+highL preferring a larger mean σ_8 than *Planck*+WP when excluding BAO data, and because σ_8 is anti-correlated with $\sum m_\nu$. When BAO is included then the mean σ_8 is comparable for *Planck*+WP and WMAP+highL. The 1D marginalised likelihoods for $\sum m_\nu$ are presented in figure III–2.9 for *Planck*+WP combined with permutations of two LSS data sets. The full set of fitted parameters are presented in the 2$^{\rm nd}$ column of table III–2.4. For the combination of *Planck*2013+WP+*All LSS (2013)*, $\sum m_\nu = (0.357 \pm 0.099)\,{\rm eV}$ compared to a 95% upper limit of $\sum m_\nu < 0.258\,{\rm eV}$ for *Planck*2013+WP+BAO. The largest change in the other fitted parameters is a $\sim 1.6\sigma$ shift in $\Omega_{\rm c}h^2$. As expected the fitted value of σ_8 shifts from 0.818 ± 0.023 to 0.749 ± 0.019 when LSS is included, but the value of $\Omega_{\rm m}$ actually increases by around 1σ, presumably since the massive neutrinos contribute to it. The value of the combined mass of neutrinos using *All LSS (Strong)* or *All LSS (Weak)* combinations with *Planck*2015+Pol included becomes $\sum m_\nu = (0.176\pm0.056)$eV or $\sum m_\nu = (0.146\pm0.057)$eV, where the significance of neutrino masses has slightly reduced from the $\sum m_\nu = (0.357\pm0.099)$eV stated above.

The significance of the increased goodness of fit can be evaluated using the likelihood ratio test. For nested models, it is particularly simple and involves

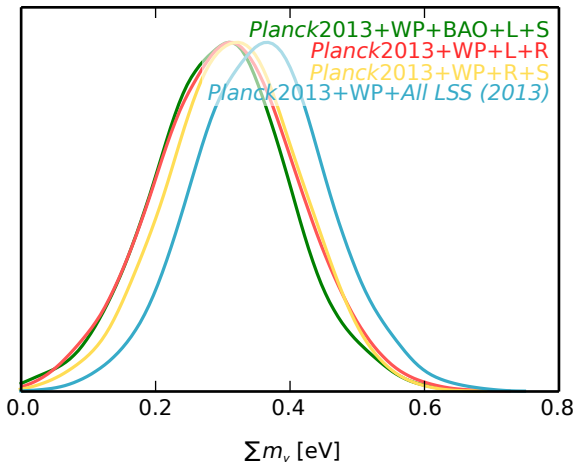


Figure III–2.9: 1D marginalised likelihoods for $\sum m_\nu$ for *Planck*2013+WP and different combinations of LSS data. The main result is *Planck*2013+WP+*All LSS (2013)* (blue), which gives $\sum m_\nu = (0.357 \pm 0.099)\,\mathrm{eV}$. The LSS data sets are labelled L for Lensing2013 + *Planck*2013 lensing, R for RSD DR12 and SZ for *Planck*2013 lensing. The combinations of two LSS probes are BAO + L + S (green), L + R (red) and R + S (yellow).

taking the difference of the maximum likelihood for the base model, in this case ΛCDM, and an extended model with $\sum m_\nu$. This can then be compared to a χ^2 distribution for one degree of freedom, with the $p-$value quantifying the probability of the simpler model being preferred over the extended model. For *Planck*2013+WP+BAO a probability of 50.2% is found but when using *Planck*2013+WP+*All LSS (2013)* the probability is only 0.35%, suggesting that the simpler model can be rejected with high probability. The high values of $\sum m_\nu$ are not favoured by the *Planck*+WP+BAO data: values of $\sum m_\nu$ for *Planck*2013+WP+*All LSS (2013)* are in tension with the upper limit from *Planck*2013+WP+BAO. As discussed in [45], this can be quantified by performing the analysis using two different neutrino masses and considering the statistics of the difference. To be concrete, in the MCMC analysis two masses, $\sum m_\nu^{\mathrm{CMB}}$ and $\sum m_\nu^{\mathrm{LSS}}$, are included and all observables (CMB power spectra, lensing convergence etc.) for each are calculated. For any CMB data the observable mass of the neutrino is $\sum m_\nu^{\mathrm{CMB}}$, and for any LSS data the observable mass of the neutrino is $\sum m_\nu^{\mathrm{LSS}}$. The other cosmological parameters are shared. The marginalised posterior $\Delta M = \sum m_\nu^{\mathrm{CMB}} - \sum m_\nu^{\mathrm{LSS}}$ can be computed to quantify at what significance this is non-zero. Performing such an analysis gives $\Delta M = 0.32 \pm 0.13$ eV, i.e. non-zero at the 2.5σ level. This quantifies the extent to which the active neutrino model is in tension with *Planck*2013 data. The

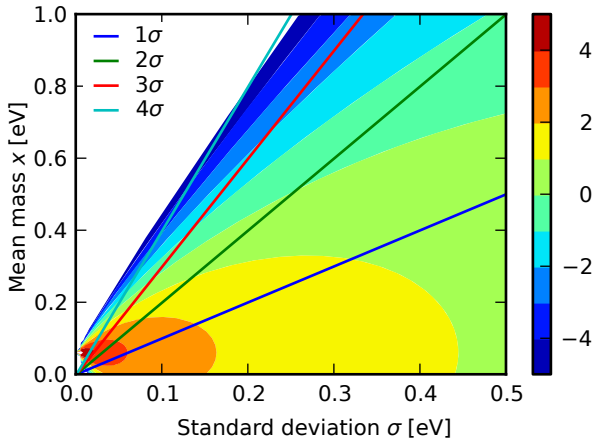


Figure III–2.10: The value of the natural logarithm of the Bayes factor for a Gaussian probability with a given mean, variance σ and fixed flat prior of [0, 3] eV. The 1, 2, 3 and 4 σ values are plotted in blue, green, red and light blue respectively.

likelihood ratio test can also be used to assess the increase in goodness of fit, finding the simpler one-neutrino model is favoured with a probability of only 0.30%.

The inclusion of active neutrinos can be seen as an addition to ΛCDM and as such model comparison can be performed using the Savage-Dickey density ratio. From equation I–3.11 in Part I–3.4, $\mathcal{M}_1$ is ΛCDM and $\mathcal{M}_2$ is ΛCDM + $\sum m_\nu$ where the additional parameter is $\psi = \sum m_\nu$. Recall that in the vanilla ΛCDM model $\sum m_\nu$ is set to 0.06eV. The normalised posterior likelihood, $P(\sum m_\nu | d, \Lambda\text{CDM} + \psi)$ is taken from the MCMC chains, where a prior range of $\sum m_\nu = [0, 3]\,\text{eV}$ is assumed. For d = *Planck*2013+WP+*All LSS (2013)*, $\ln(\Theta_\text{B}) = -1.8$ implying that a model including active neutrinos is preferred over plain ΛCDM by odds of around 6 : 1. This represents reasonably strong evidence on the Jeffreys scale [196]. If instead of the prior $[0, 3]\,\text{eV}$, which is not unreasonable but is also not compelling, $[0, 1]\,\text{eV}$ or $[0, 10]\,\text{eV}$, then $\ln(\Theta_\text{B}) = -2.9$ and -0.6 respectively. These equate to odds from 18:1 to 2:1. Figure III–2.10 shows the value of the logarithm of the Bayes factor when a Gaussian probability is assumed with some flat prior,

$$\ln(\Theta_\text{B}) \quad = \quad \ln\left(\frac{e^{(x_0-x)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \times \frac{1}{p(x)}\right), \tag{III–2.31}$$

where σ is the standard deviation, x is the mean mass, x_0 is the fiducial mass and $p(x)$ is the uniform prior range normalised to 1. For ΛCDM + $\sum m_\nu$ then

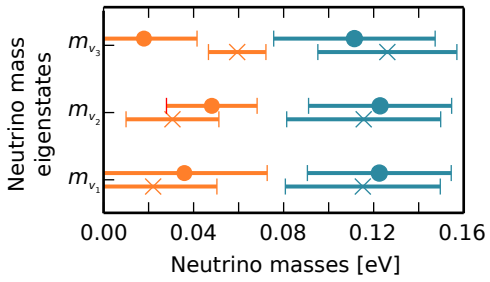


Figure III–2.11: Masses of the neutrino eigenstates using *Planck*2013+WP+BAO (orange) and *Planck*2013+WP+*All LSS (2013)* (blue), together with results from oscillation experiments [78]. These allow for two solutions, called normal (crosses) and inverted (circles) hierarchies. When the higher $\sum m_\nu$ value from *Planck*2013+WP+*All LSS (2013)* is used then the eigenstates become degenerate.

$x = 0.06$eV and with the prior $[0, 3]$ eV, $p(x) = 1/3$. For $\sum m_\nu = 0.357 \pm 0.099$ obtained from *Planck*2013+WP+*All LSS (2013)*, the value of the Bayes factor is $\ln(\Theta_{\rm B}) \sim -2$ and has between 3-4σ significance (red and light blue lines in figure III–2.10). As such, not only are the Bayesian statistics showing reasonably strong odds, but so is the statistical likelihood. If the prior range is changed to some larger value, the Bayes factor becomes larger and starts supporting no change from standard ΛCDM, but the statistical likelihood would continue to show that active neutrinos are significant within the 3-4σ range.

Cosmological limits on $\sum m_\nu$ for active neutrinos are important since they can be combined with square differences between the neutrino masses obtained from atmospheric, $\Delta m^2_{\rm A}$, and solar neutrino measurements, $\Delta m^2_{\rm S}$, in order to calculate the masses of the individual eigenstates, $(m_{\nu_1}, m_{\nu_2}, m_{\nu_3})$ [274]. An important question to answer is whether the masses have a "normal", "inverted" or "degenerate" hierarchy. From [78] the difference between $m^2_{\nu_3}$ and $m^2_{\nu_1}$ is given by $\Delta m^2_{\rm A} = 2.4^{+0.1}_{-0.1} \times 10^{-3}$eV^2 (i.e. an uncertainty of $\sim 4\%$) and the difference between $m^2_{\nu_2}$ and $m^2_{\nu_1}$ is given by $\Delta m^2_{\rm S} = 7.5^{+0.3}_{-0.2} \times 10^{-5}$eV^2 (i.e. an uncertainty of $\sim 3\%$). For normal hierarchy, $m^2_{\nu_1}$ and $m^2_{\nu_2}$ are much less than $m^2_{\nu_3}$ and the mass eigenstates are given by $m^2_{\nu_2} = m^2_{\nu_1} + \Delta m^2_{\rm S}$ and $m^2_{\nu_3} = m^2_{\nu_1} + \Delta m^2_{\rm A}$ with m_{ν_1} being the lowest mass eigenstate, and in the inverted hierarchy, $m^2_{\nu_3}$ is the lowest mass eigenstate with a mass much less than $m^2_{\nu_1} = m^2_{\nu_3} + \Delta m^2_{\rm A} - \Delta m^2_{\rm S}$ and $m^2_{\nu_2} = m^2_{\nu_3} + \Delta m^2_{\rm A}$. In figure III–2.11 the individual neutrino masses are shown for $\sum m_\nu$ given the limits from *Planck*2013+WP+BAO and *Planck*2013+WP+*All LSS (2013)*, i.e. smallest and largest $\sum m_\nu$ values respectively. The normal hierarchy equations are plot-

ted with crosses and the inverted with circles. *Planck*2013+WP+BAO predicts the mass of three active neutrinos to be $< 0.258\,\mathrm{eV}$ with m_{ν_3} greater than m_{ν_1} and m_{ν_2} indicating that, if this limit is correct, then a "normal" hierarchy is preferred, although "inverted" or "degenerate" hierarchy are by no means excluded. The preferred masses are $m_{\nu_1} = (0.022 \pm 0.028)\,\mathrm{eV}$, $m_{\nu_2} = (0.031 \pm 0.021)\,\mathrm{eV}$ and $m_{\nu_3} = (0.059 \pm 0.013)\,\mathrm{eV}$. The same is true when looking at the inverted hierarchy equation, where clear preference is seen for "inverted" whilst "normal" and "degenerate" hierarchies are not ruled out. The masses in this case are $m_{\nu_1} = (0.036 \pm 0.028)\,\mathrm{eV}$, $m_{\nu_2} = (0.048 \pm 0.020)\,\mathrm{eV}$ and $m_{\nu_3} = (0.018 \pm 0.024)\,\mathrm{eV}$. Interestingly, when looking at the inverted hierarchy equation, m_{ν_1} is unbounded from below like m_{ν_3} is whereas m_{ν_3} is statistically unlikely to have zero mass in the normal hierarchy case, perhaps suggesting slightly more preference for normal hierarchy. *Planck*2013+WP+*All LSS (2013)* constrains $\sum m_\nu = (0.357 \pm 0.099)\mathrm{eV}$ which is more consistent with a degenerate hierarchy for both the normal and inverted equations. The preferred masses in this case are $m_{\nu_1} = (0.115 \pm 0.034)\,\mathrm{eV}$, $m_{\nu_2} = (0.116 \pm 0.034)\,\mathrm{eV}$ and $m_{\nu_3} = (0.126 \pm 0.031)\,\mathrm{eV}$ or $m_{\nu_1} = (0.123 \pm 0.032)\,\mathrm{eV}$, $m_{\nu_2} = (0.123 \pm 0.032)\,\mathrm{eV}$ and $m_{\nu_3} = (0.111 \pm 0.036)\,\mathrm{eV}$ for normal and inverted equations respectively. Each indicate more than 3σ preference for non-zero neutrino mass. There is strong correlation between the probability distribution for each eigenstate. The results using *Planck*+WP+*All LSS (2013)* reveal that, regardless of hierarchy equation, degenerate hierarchy seems to be preferred rather strongly in comparison to CMB data alone which is compatible with all three.

Sterile neutrinos

There are a host of anomalies from short baseline neutrino oscillation experiments which may be solved by the addition of a sterile neutrino. Firstly, the LSND experiment [20] observes an excess of $\bar{\nu}_e$ candidates, suggesting the oscillation $\bar{\nu}_\mu \to \nu_s \to \bar{\nu}_e$ where the mass of the sterile neutrino is constrained to $\sim$ 1eV by the KARMEN [29] and Bugey [102] experiments. The MiniBooNE experiment [21], as well as testing the LSND signal, also detects an excess of ν_e from the neutrino mode rather than anti-neutrino mode. Although the neutrino mode does not completely agree with the expected sterile neutrino signal, there are several explanations due to the method of detection [209]. Reactor anomalies detect a 6% lower rate of electron anti-neutrinos than is expected, which can be interpreted as neutrino oscillations with a 1eV sterile neutrino [261]. Lower event rates of $\nu_e + {}^{71}\mathrm{Ga} \to {}^{71}\mathrm{Ge} + e^-$ than expected can also be explained by 1eV sterile neutrino oscillations, solving what is known as the Gallium anomaly [144].

Joint analysis using cosmological and short baseline data has been carried out for models with both one and two added sterile neutrinos [28]. The addition of short baseline data in the form of priors on the cosmological data lowers the mass of the sterile neutrino, in a single sterile neutrino model, to the $m_{\rm sterile} \sim 1$eV range at high significance. A model with two added sterile neutrinos is generally not allowed since this leads to a universe with too much radiation [232, pg. 163]. When considering two mixing angles between active and sterile neutrinos a small allowed parameter space is left putting tension between short baseline neutrino oscillation experiments and cosmological bounds. This can be rectified by including new parameters in the cosmological model, such as large asymmetries between neutrinos and antineutrinos [263]. This problem is not considered here, since the more general addition of sterile neutrinos is concentrated on, rather than a specific model.

The standard approach to modelling sterile neutrinos is to introduce two new parameters into the fitting process. The first is an effective neutrino mass, $m^{\rm eff}_{\rm sterile}$, and the second is the change in the effective number of degrees of freedom, $\Delta N_{\rm eff}$, such that $N_{\rm eff} = 3.046 + \Delta N_{\rm eff}$. In this case the active neutrinos are modelled as one massive neutrino with $m_\nu = 0.06\,{\rm eV}$ and two massless neutrinos, which would accurately model a normal hierarchy with $m_1 < m_2 \ll m_3$. The cosmological results are not very sensitive to this assumption on the structure of the neutrino active sector. The density of sterile neutrinos is given by

$$\Omega_{\rm sterile} h^2 \quad = \quad \frac{m^{\rm eff}_{\rm sterile}}{94\,{\rm eV}}\,. \tag{III–2.32}$$

In the MCMC analysis, a prior on $m^{\rm eff}_{\rm sterile}/\Delta N_{\rm eff} < 10\,{\rm eV}$ is imposed, as used in the *Planck*2013 analysis [9], since sterile neutrinos with large effective masses become degenerate with CDM.

This parameterisation encompasses a wide of range of possible models for sterile neutrinos, which are typically formed in the early Universe by oscillations. Two possible scenarios that have been widely discussed in the literature are:

"On resonance" oscillations in the Dodelson-Widrow model [110]. In this case the sterile neutrinos have the same temperature as their active counterparts and are formed via oscillations when there is no lepton asymmetry and the mixing angle is small. Under the assumption that neutrino decoupling is instantaneous, the distribution function for the neutrinos is

$$f_{\rm DW}(p) \quad = \quad \frac{\Delta N_{\rm eff}}{\exp[p/T_\nu] + 1}\,. \tag{III–2.33}$$

In this case $\Delta N_{\rm eff}$ is a constant, which is not necessarily an integer, and the true mass is given by $m_{\rm sterile} = m^{\rm eff}_{\rm sterile}/\Delta N_{\rm eff}$. This means that the distribution function is suppressed with respect to one of the active neutrinos.

"Off resonance" oscillations leading to a thermal scenario [332]. The distribution function is standard Fermi-Dirac with thermal temperature $T_{\rm sterile}$, given by

$$T_{\rm sterile} \quad = \quad \Delta N_{\rm eff}^{1/4} T_\nu \,, \tag{III–2.34}$$

where T_ν is the thermal temperature of the active neutrinos. $\Delta N_{\rm eff}$, again not necessarily an integer, quantifies the level of thermalisation with $\Delta N_{\rm eff} = 1$ corresponding to the complete thermalisation, for one species of sterile neutrino, with $T_{\rm sterile} = T_\nu$. The thermal mass of the neutrinos is given by

$$m_{\rm sterile} \quad = \quad \Delta N_{\rm eff}^{-3/4} m^{\rm eff}_{\rm sterile} \,. \tag{III–2.35}$$

The detailed analysis of a wide range of LSS data combined with CMB data from *Planck*2013+WP or WMAP+highL on active neutrinos showed that combinations of the CMB and two or more LSS data sets lead to consistent conclusions. Similar is true for results from the sterile neutrino case. For this reason, the results are restricted to a limited range of possibilities.

Figure III–2.12 illustrates the impact of including $m^{\rm eff}_{\rm sterile}$ and $\Delta N_{\rm eff}$ as parameters on the constraints in the $\sigma_8 - \Omega_{\rm m}$ plane for various data combinations. Unlike in the case of active neutrinos, in the left subplot there is clear evidence that inclusion of sterile neutrinos can ameliorate, at least partially, the discrepancy between the *Planck*2013+WP+BAO and *All LSS (2013)* parameter distributions. Figure III–2.13 presents the results of joint CMB and LSS analyses. There is a consistent picture for a range of combinations of these data suggesting a non-zero value for $m^{\rm eff}_{\rm sterile}$. Using *Planck*2013+WP+*All LSS (2013)* it can be seen that $m^{\rm eff}_{\rm sterile} = (0.67 \pm 0.18)\,{\rm eV}$ and $\Delta N_{\rm eff} = 0.32 \pm 0.20$ - marginalised parameters are presented in Table III–2.5.

Comparing the newer data sets of *Planck*2015+Pol+BAO, *All LSS (Weak)* and *(Strong)*, sterile neutrinos fare a little better than their active counterparts in reducing the tension. When $m^{\rm eff}_{\rm sterile}$ and $\Delta N_{\rm eff}$ are included in the quantification analysis $C = 0.891$ (1.60σ) and $C = 0.652$ (0.94σ) comparing *Planck*2015+Pol+BAO to *All LSS (Strong)* and *All LSS (Weak)* respectively. This is in good agreement with the visual inspection of $\Omega_{\rm m} - \sigma_8$ contours. Due to the high dimensionality of this problem each bin in the histogram for $m^{\rm eff}_{\rm sterile}$ and $\Delta N_{\rm eff}$ is computed separately, written to disk and then

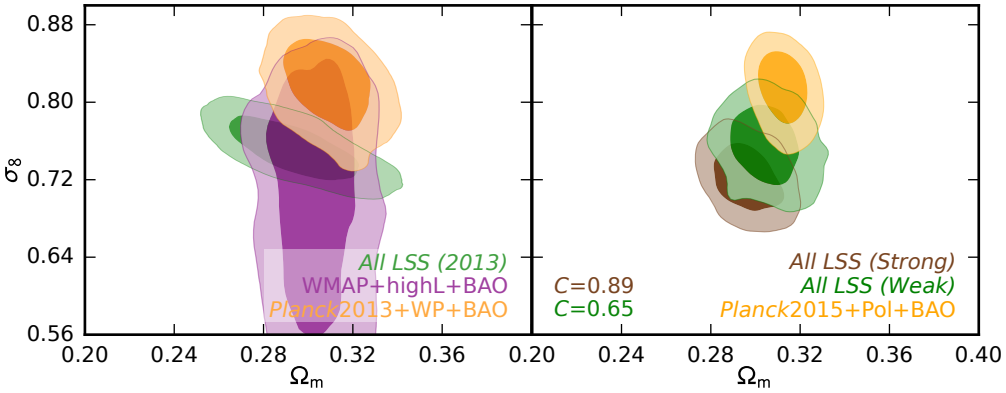


Figure III–2.12: 1σ and 2σ constraints in the $\Omega_{\rm m} - \sigma_8$ plane for ΛCDM+$m^{\rm eff}_{\rm sterile}$+$\Delta N_{\rm eff}$ from *Planck*2013+WP+BAO (orange) and WMAP+highL+BAO (purple) on the left and *Planck*2015+Pol+BAO (light orange) on the right and from combining each of the LSS data sets, *All LSS (2013)* (lighter green) on the left and *All LSS (Strong)* (brown) and *All LSS (Weak)* (darker green) on the right. The five parameter ΛCDM difference vector with *Planck*2015+Pol+BAO is quoted for both sets of constraints in the bottom-left corner of the right subplot.

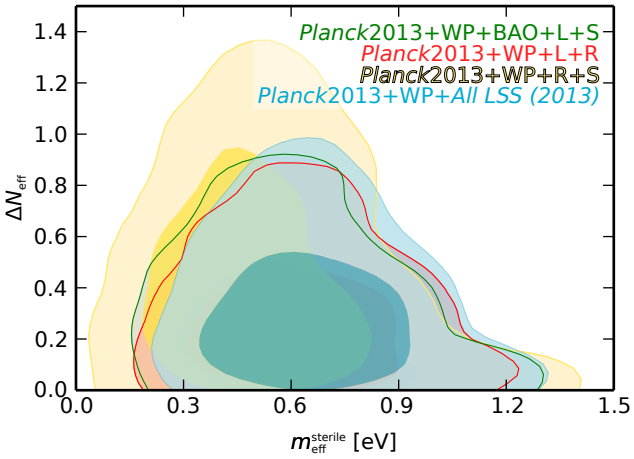


Figure III–2.13: 2D Likelihood plots for $m^{\rm sterile}_{\rm eff}$ and $\Delta N_{\rm eff}$ for *Planck*2013+WP combined with different LSS data combinations. The main result is *Planck*2013+WP+*All LSS (2013)* (blue). The LSS data sets are labelled L for Lensing2013 + *Planck*2013 lensing, R for RSD DR12 and SZ for *Planck*2013 lensing. The combinations of two LSS probes: BAO + L + S (green), L + R (red) and R + S (yellow).

Parameter	*Planck*2013+WP+BAO	*Planck*2013+WP+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02237±0.00029	0.02250±0.00028
$\Omega_{\rm c}h^2$	0.1250±0.0050	0.1180±0.0045
$100\theta_{\rm MC}$	1.04067±0.00073	1.04122±0.00063
$\tau_{\rm R}$	0.096±0.014	0.096±0.015
$n_{\rm S}$	0.977±0.010	0.969±0.010
$\log(10^{10}A_{\rm s})$	3.115±0.029	3.105±0.030
$m^{\rm eff}_{\rm sterile}$[eV]	<0.48	0.66±0.18
$\Delta N_{\rm eff}$	0.47±0.27	0.32±0.21
H_0	69.79±1.71	68.00±1.11
$\Omega_{\rm m}$	0.308±0.012	0.321±0.013
σ_8	0.817±0.030	0.736±0.017
$z_{\rm re}$	11.77±1.18	11.66±1.25

Table III–2.5: Marginalised parameter table when ΛCDM is extended with sterile neutrinos. The data combinations are *Planck*2013+WP+BAO in the first column and *Planck*2013+WP+*All LSS (2013)* in the second.

analysed from the disk. This increases computation times significantly, especially when testing for a range of bin sizes and amounts of Gaussian smoothing. The values of $m^{\rm eff}_{\rm sterile} = (0.470 \pm 0.227)$eV and $\Delta N_{\rm eff} = 0.093 \pm 0.057$ or $m^{\rm eff}_{\rm sterile} = (0.234\pm0.115)$eV and $\Delta N_{\rm eff} = 0.116\pm0.059$ are obtained by combining *All LSS (Strong)* or *All LSS (Weak)* combinations with *Planck*2015+Pol. These constraints are similar to the values expected from *Planck*2015+Pol+BAO, although with peaks in their respective distributions.

Since the *Planck*2013+WP+*All LSS (2013)* results show a significant preference for the sterile neutrino model using the joint likelihood it is again useful to quantify using the likelihood ratio test, with two extra degrees of freedom, $m^{\rm eff}_{\rm sterile}$ and $\Delta N_{\rm eff}$. The probability of the simpler ΛCDM model being preferred is 89.3% using *Planck*2013+WP+BAO, but 0.00% when using *Planck*2013+WP+*All LSS (2013)*.

It was pointed out in [45; 382] that the extra degree of freedom due to $\Delta N_{\rm eff}$ allowed for a best-fitting value of H_0 more compatible with low redshift measurements, for example, using Cepheids [301]. This is due to the degeneracy between $\Delta N_{\rm eff}$ and H_0. However, with the inclusion of RSD an even lower value of σ_8 is preferred (see figures III–2.2 and III–2.3). Due to the $\sigma_8 - H_0$ degeneracy the result is a lower value of H_0 more closely aligned with the *Planck*2013+WP+BAO

value.

As with active neutrinos, the improvement in the likelihood when including the extra parameters comes at a price; the *Planck*2013 component of the likelihood is increased by $\Delta\chi^2 \approx 4$. This is less problematic than in the case of active neutrinos since the extra freedom from $\Delta N_{\rm eff}$ allows a better fit to the CMB, but it is still unsatisfactory. Performing a similar analysis to quantify the residual level of tension CMB parameters, $m_{\rm sterile}^{\rm eff,CMB}$ and $\Delta N_{\rm eff}^{\rm CMB}$, and LSS parameters $m_{\rm sterile}^{\rm eff,LSS}$ and $\Delta N_{\rm eff}^{\rm LSS}$ are defined, with all other parameters shared. As before, each CMB or LSS observable uses the relevant neutrino parameters. The joint marginalised distribution of $m_{\rm sterile}^{\rm eff,LSS} - m_{\rm sterile}^{\rm eff,CMB}$ and $\Delta N_{\rm eff}^{\rm LSS} - \Delta N_{\rm eff}^{\rm CMB}$ is then constructed, and it is found that both parameters are non-zero at the 1.6σ level. In this case the likelihood ratio test favours the one sterile neutrino model with a probability of 8.60%, again indicating less internal disagreement between the CMB and LSS values than the active case.

The Savage-Dickey density ratio can again be calculated, where the extension parameters are $\psi = (m_{\rm eff}^{\rm sterile}, \Delta N_{\rm eff})$. The unextended posterior likelihood, $P(\psi|d, \Lambda{\rm CDM} + \psi)$ is obtained from MCMC by setting $m_{\rm eff}^{\rm sterile} = 0\,{\rm eV}$ and $N_{\rm eff} = 3.046$. The priors used are $m_{\rm eff}^{\rm sterile} = [0, 3]\,{\rm eV}$ and $N_{\rm eff} = [3.046, 10]$. Again using $d =$ *Planck*2013+WP+*All LSS (2013)* then $\ln(\Theta_{\rm B}) = -2.67$, which strongly supports the addition of sterile neutrinos over *vanilla* ΛCDM. This is in contrast to the values presented in [230]. For a similar data combination they find that $\ln(\Theta_{\rm B}) \approx 1$. Since they use a similar prior range, $m_{\rm eff}^{\rm sterile} = [0, 3]\,{\rm eV}$, it appears that the discrepancy is due to the posterior likelihood for $m_{\rm eff}^{\rm sterile}$. Although no specific numbers are presented, it is clear from the right-hand panel of figure 1 in [230] that their constraint is much weaker than the $m_{\rm eff}^{\rm sterile} = 0.67 \pm 0.18$ we report here. The values in this Thesis are compatible with those presented in [45; 156; 382] all of which suggest $\sim 4\sigma$ preferences for sterile neutrinos, albeit for slightly different data combinations. It is clear from the figure in [230] that a 4σ detection with a central value of $0.2 - 0.4$, as reported, has a Bayes factor that should be somewhere in the range -2 to -4, compatible with the values presented here, and contrary to [230].

2.4.2 Alternative explanations

It has been noted in Part III–2.4.1 that tensions between *All LSS (2013)* and *Planck*2013+WP+BAO measurements of the CMB are reduced with the addition of neutrinos, although this is not seen to as great an extent when considering *All LSS (Weak)* and *All LSS (Strong)* with *Planck*2015+Pol+BAO. Although

the discordance is partially reduced in the former case, there may be better modifications to the standard cosmological model that might accommodate the two types of data. Some possible explanations are presented in this section.

Modifications to the primordial power spectrum

The inclusion of massive neutrinos reduces the amount of small-scale power relative to large-scales in the observed matter power spectrum. One obvious possibility that needs to be considered is whether such an effect can be created using a simple modification to the primordial power spectrum of adiabatic perturbations, $P_i(k)$. Within the ΛCDM model this is $P_i(k) = P_{\Lambda\mathrm{CDM}} = A_\mathrm{s}(k/k_\mathrm{pivot})^{n_\mathrm{s}}$ where k_pivot is fixed to be $k_\mathrm{pivot} = 0.05\,\mathrm{Mpc}^{-1}$. Similar endeavours have been made previously in the literature, as in [126], where the primordial power spectrum is modified via the application of a set of "top-hat" steps or a "sawtooth" shaped function to the original power spectrum. This was done as an attempt to explain specific features which could not easily be explained by a power law, $P_i(k) \propto k^n$, such as a bump-like feature in the CMB at $k \sim 0.004\,h\,\mathrm{Mpc}^{-1}$ [150; 159], a step-like feature between $k \sim 0.06 - 0.6\,h\,\mathrm{Mpc}^{-1}$ [43] and a dip at $k \sim 0.1\,h\,\mathrm{Mpc}^{-1}$ [149]. Inflationary features can be included by modifying the primordial power spectrum. Step-like features, in particular, can be caused by interacting scalar fields which, in turn, cause localised oscillations in the CMB which can provide a better fit to data than a featureless power spectrum [50; 51; 201].

Modifications to the primordial power spectrum that can mimic the effects of including massive neutrinos are considered here. In particular, the specific form of the modification is

$$P_i(k) \quad = \quad \left[1 - \frac{\alpha}{2}\left(1 + \tanh\left(\frac{\ln \beta k}{\ln \delta}\right)\right)\right] P_{\Lambda\mathrm{CDM}}(k)\,, \qquad \text{(III–2.36)}$$

where α determines the magnitude of the overall suppression and β and δ control the position and rate of the turn over, respectively. For example, $\alpha = 0.14$, $\beta = 20\,\mathrm{Mpc}$ and $\delta = 5$ mimics the matter power spectrum of an active neutrino model with $\sum m_\nu = 0.3\,\mathrm{eV}$. Note that having the same observed matter power spectrum does not imply that the CMB power spectrum will be the same. The matter power spectrum and CMB temperature power spectrum are plotted in figures III–2.14 and III–2.15 for a range of different modified $P_i(k)$ parameter values, as well as for a range of different active and sterile neutrino parameters. A similar form of modification to the primordial power spectrum was proposed in [165], which examined how binning the primordial spectrum can produce features similar to Starobinsky inflation [345]. They found that a sharp transition

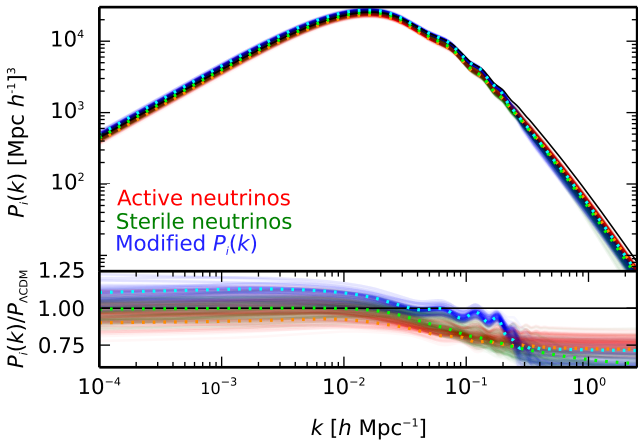


Figure III–2.14: Predicted matter power spectra for models colour coded by their fits to likelihoods: (i) active neutrinos (red); (ii) sterile neutrinos (green); (iii) modified $P(k)$ (blue). The overlayed dotted lines show the best fit results for each model.

is equally as probable as a smooth transition, since δ is unconstrained from below.

After marginalisation using *Planck*2013+WP+*All LSS (2013)* $\alpha = 0.32 \pm 0.11$, $\beta = 5.96 \pm 0.70$ Mpc and $\delta = 1.24 \pm 0.11$ with best-fits $\alpha = 0.20$, $\beta = 6.76$ Mpc and $\delta = 1.12$. The combined likelihood is improved in comparison to the addition of sterile and active neutrinos by $\Delta\chi^2 \approx 1$ and by $\Delta\chi^2 \approx 11$ respectively, which can be attributed to a better fit to *Planck*2013 temperature data. The *Planck* data has a "dip" at around $\ell = 1800$ which corresponds with a residual systematic feature due to incomplete 4K line removal [9]. It is possible the modified $P_i(k)$ model fits this feature better, as seen by the reduction in power around $\ell = 1800$ in the figure III–2.15. The modified $P_i(k)$ model fits LSS data as well as the neutrino models.

Given the fact there is a reduction in power for $\ell \gtrsim 2000$, *Planck*2013+WP+*All LSS (2013)*+highL has also been used, where highL is the reduced ACT+SPT data used in the *Planck*2013 analysis. The shape of the hyperbolic tangent function remains similar with highL data, $\beta = 6.00 \pm 2.51$Mpc and $\delta = 0.92 \pm 0.48$, but the amplitude, $\alpha = 0.111 \pm 0.083$, is tightly constrained. Therefore, it appears that this model can be excluded on the basis of α being consistent with zero. The fit is also worsened in comparison to the previous case, where highL is not included, with a $\Delta\chi^2 \approx 10$.

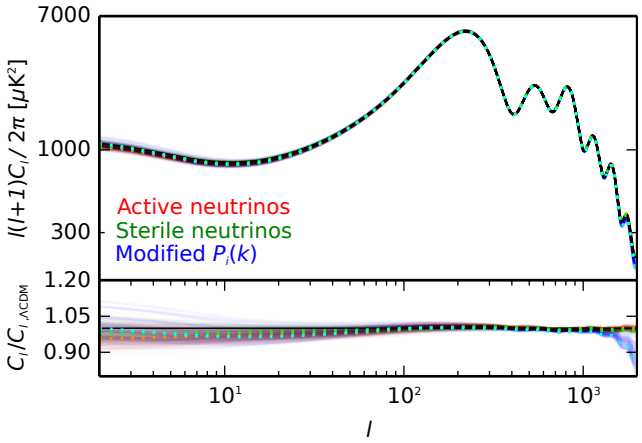


Figure III–2.15: CMB temperature power spectra for models colour coded by their fits to likelihoods: (i) active neutrinos (red); (ii) sterile neutrinos (green); (iii) modified $P(k)$ (blue). The overlayed dotted lines show the best fit results for each model.

Varying the lensing parameter, $A_{\rm L}$

Weak lensing has two effects on the CMB: the first is an additional contribution to the angular power spectrum, the second is a non-Gaussian effect that has an impact on the higher-order moments. This latter effect is used in the reconstruction that had already been part of this analysis. One odd effect that has been documented in the *Planck*2013 analysis (the result being more significant when including highL data) is that, when one adds a phenomenological parameter such that $C_\ell^\psi \rightarrow A_{\rm L} C_\ell^\psi$ [76], with $A_{\rm L} = 0$ corresponding to an unlensed model and $A_{\rm L} = 1$ the physical result, one finds $A_{\rm L} = 1.23 \pm 0.11$ [9] (*Planck*2013+WP+highL) which is $\sim 2\sigma$ away from the expected value. The reasoning behind this high value of $A_{\rm L}$ being favoured is a mystery.

The impact of varying of $A_{\rm L}$ on the models considered in tables III–2.4 and III–2.5 and the equivalent results are presented in table III–2.6 and III–2.7. It is seen that larger values of $\sum m_\nu$ and $m^{\rm eff}_{\rm sterile}$ are allowed for *Planck*+WP due to the degeneracy with $A_{\rm L}$. However, when including *All LSS (2013)*, the fit to the *Planck*2013 component of the likelihood is still degraded, although this is less severe for active neutrinos than when $A_{\rm L} = 1$. With *All LSS*, the significance of the active neutrino result increases to $\sum m_\nu = (0.420 \pm 0.097)\,{\rm eV}$, but for a sterile neutrino the mass decreases, such that $m^{\rm eff}_{\rm sterile} = (0.56 \pm 0.15)\,{\rm eV}$. The reason for this is the correlation between the effective sterile neutrino mass and several of the observable parameters. For example, H_0 anti-correlates with $m^{\rm eff}_{\rm sterile}$ and

Parameter	*Planck*2013+WP+BAO	*Planck*2013+WP+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02247±0.00030	0.02244±0.00024
$\Omega_{\rm c}h^2$	0.1160±0.0023	0.1141±0.00142
$100\theta_{\rm MC}$	1.04174±0.00063	1.04179±0.00055
$\tau_{\rm R}$	0.087±0.013	0.089±0.012
$n_{\rm S}$	0.9696±0.0070	0.9718±0.0056
$\log(10^{10}A_{\rm s})$	3.075±0.025	3.073±0.023
$A_{\rm L}$	1.28±0.13	1.182±0.055
$\sum m_\nu$[eV]	<0.47	0.420±0.097
H_0	67.75±1.04	66.58±1.14
$\Omega_{\rm m}$	0.307±0.013	0.319±0.014
σ_8	0.777±0.038	0.725±0.019
$z_{\rm re}$	10.65±1.08	10.86±1.05

Table III–2.6: Marginalised parameter table for the same model presented in table III–2.4 (ΛCDM+$\sum m_\nu$) when the amplitude of the lensing contribution the CMB temperature power spectrum, $A_{\rm L}$, is allowed to vary.

since H_0 is larger when $A_{\rm L}$ is allowed to vary, this would suggest a lower sterile neutrino mass. $\Omega_{\rm m}$ and $z_{\rm re}$ are both correlated with $m^{\rm eff}_{\rm sterile}$ and since both $\Omega_{\rm m}$ and $z_{\rm rei}$ have lower values when $A_{\rm L}$ is varied, then this also corresponds with $m^{\rm eff}_{\rm sterile}$ being smaller.

Ignore WMAP polarisation data

The measurement of the E-mode polarisation on large-scales by WMAP is crucial in all the previous analyses containing WP. The CMB temperature anisotropies constrain the parameter combination $A_{\rm s}e^{-2\tau_{\rm R}}$ in the absence of the ISW effect and this requires a measurement of polarisation on large scales to infer $\tau_{\rm R}$ and hence allow $A_{\rm s}$ to be deduced independently. The small-scale amplitude σ_8 is a derived parameter and is sensitive to all the cosmological parameters, but it is proportional to the square root of the amplitude of scalar perturbations, $\sqrt{A_{\rm s}}$. If $\tau_{\rm R}$ were lower than the $\tau_{\rm R} = 0.091\pm0.013$ as required by *Planck*2013+WP+BAO then σ_8 would reduce $\propto e^{\tau_{\rm R}}$. In particular a reduction of σ_8 from ≈ 0.83 as suggested by CMB measurements to ≈ 0.78, which is closer to the value preferred by the LSS measurements, would require $\tau_{\rm R}$ to reduce from ≈ 0.09 to ≈ 0.05. Of course, this would require the WMAP polarisation measurement on large scales to have been misinterpreted. However, this is the regime where instrumental systematics and foreground subtraction are most difficult and therefore it seems at least sensible to consider such a possibility.

Parameter	*Planck*2013+WP+BAO	*Planck*2013+WP+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02286±0.00037	0.02277±0.00030
$\Omega_{\rm c}h^2$	0.1232±0.0052	0.1213±0.0043
$100\theta_{\rm MC}$	1.0410±0.00075	1.0410±0.00066
$\tau_{\rm R}$	0.092±0.014	0.090±0.013
$n_{\rm S}$	0.985±0.012	0.982±0.012
$\log(10^{10}A_{\rm s})$	3.103±0.030	3.095±0.026
$A_{\rm L}$	1.29±0.13	1.137±0.063
$m^{\rm eff}_{\rm sterile}$[eV]	<0.82	0.56±0.15
$\Delta N_{\rm eff}$	0.55±0.27	0.54±0.26
H_0	70.59±1.71	69.66±0.15
$\Omega_{\rm m}$	0.302±0.012	0.311±0.013
σ_8	0.771±0.044	0.732±0.017
$z_{\rm re}$	11.23±1.21	11.12±1.10

Table III–2.7: Marginalised parameter table for the same model presented in table III–2.5 (ΛCDM+$m^{\rm eff}_{\rm sterile}$ + $\Delta N_{\rm eff}$) when the amplitude of the lensing contribution the CMB temperature power spectrum, $A_{\rm L}$, is allowed to vary.

In order to illustrate the point that the LSS measurements can be used to fix $\tau_{\rm R}$ in the absence of the a large scale polarisation measurement, WP can be removed from the likelihood and the standard ΛCDM model fitted to the *Planck*2013+*All LSS (2013)* data. The results from doing this are presented in the final column of table III–2.8 and can be compared to using only *Planck*2013+WP+BAO data in the first column. The marginalised distributions for $\tau_{\rm R}$ are presented in figure III–2.16. For *Planck*2013+WP+BAO there is a narrow range of values of $\tau_{\rm R}$, but for *Planck*2013+*All LSS (2013)* the likelihood distribution for $\tau_{\rm R}$ is much wider and $\tau_{\rm R} = 0.049 \pm 0.021$. As with the both the active and sterile neutrino case, the improved fit to the LSS data leads to a degradation in the fit to the *Planck*2013 temperature data quantified by $\Delta\chi^2 \approx 6$. It is clear that the *Planck*2013+WP+BAO data has a preference for $\tau_{\rm R} \sim 0.1$ but this is not sufficiently strong to prevent the LSS dragging it to lower values in order reduce σ_8.

The lower value of $\tau_{\rm R}$ corresponds to a redshift reionisation of $z_{\rm re} = 6.91 \pm 2.20$ for *Planck*2013+*All LSS (2013)*, much lower than the value of $z_{\rm re} \approx 11$ preferred by *Planck*2013+WP+BAO. However, such values might be considered desirable in the context of astrophysical constraints on reionisation. Light from quasars show Lyman-α absorption, due to neutral hydrogen in the intergalactic

Parameter	*Planck*2013+WP+BAO	*Planck*2013+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02210±0.00025	0.02224±0.00026
$\Omega_{\rm c}h^2$	0.1187±0.0018	0.1166±0.0017
$100\theta_{\rm MC}$	1.04137±0.00058	1.04143±0.00058
$\tau_{\rm R}$	0.091±0.013	0.049±0.021
$n_{\rm S}$	0.9618±0.0058	0.9635±0.0061
$\log(10^{10}A_{\rm s})$	3.088±0.025	3.000±0.039
H_0	67.73±0.80	68.64±0.80
$\Omega_{\rm m}$	0.309±0.011	0.296±0.010
σ_8	0.825±0.012	0.783±0.012
$z_{\rm re}$	11.13±1.08	6.91±2.20

Table III–2.8: Marginalised parameter table when $\tau_{\rm R}$ is allowed to vary. The data combinations are *Planck*2013+BAO in the first column and *Planck*2013+*All LSS (2013)* in the second.

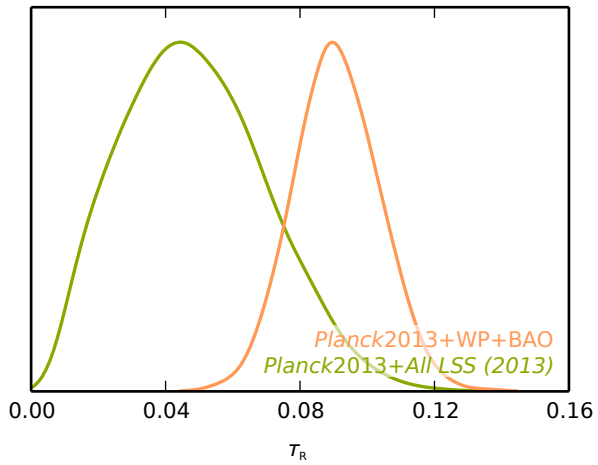


Figure III–2.16: Marginalised likelihood distributions for $\tau_{\rm R}$ within the standard ΛCDM model, for *Planck*2013+*All LSS (2013)* (green) and *Planck*2013+WP+BAO (orange). Using WP to constrain $\tau_{\rm R}$ leads to a narrow distribution of values centred on 0.09 whereas using the LSS data favours a much lower value, albeit with a wider distribution.

Parameter	*Planck*2013+WP+BAO	*Planck*2013+*All LSS (2013)*
$\Omega_{\rm b}h^2$	0.02235±0.00028	0.02236±0.00025
$\Omega_{\rm c}h^2$	0.1177±0.0019	0.1160±0.0014
$100\theta_{\rm MC}$	1.04159±0.00059	1.04160±0.00057
$\tau_{\rm R}$	0.086±0.013	<0.047
$n_{\rm S}$	0.9657±0.0061	0.9660±0.0054
$\log(10^{10}A_{\rm S})$	3.077±0.025	2.955±0.023
$A_{\rm L}$	1.19±0.10	1.202±0.055
H_0	68.37±0.89	69.02±0.67
$\Omega_{\rm m}$	0.301±0.011	0.292±0.008
σ_8	0.817±0.012	0.7605±0.0086
$z_{\rm re}$	10.60±1.11	3.99±1.47

Table III–2.9: Marginalised parameter table for the same model presented in table III–2.8 (ΛCDM+$\tau_{\rm R}$) when the amplitude of the lensing contribution the CMB temperature power spectrum, $A_{\rm L}$, is allowed to vary.

medium, at a range of frequencies depending on the redshift. The Lyman-α forest is more greatly populated for larger redshift quasars, but at $z \sim 6$ all electromagnetic radiation flux below the Lyman-α forest drops to zero, known as the Gunn-Peterson trough [151]. This effect is due to a large fraction of neutral hydrogen, hence indicating the boundary at the end of reionisation. A complete Gunn-Peterson trough can be seen in objects such as a quasar at $z = 6.28$ [49]. Objects at even lower redshifts, $z \sim 5.5$, are seen to have partial Gunn-Peterson troughs suggesting that the end of reionisation was patchy [108]. These redshifts are much lower than the predicted redshifts for the beginning of reionisation from WMAP polarisation data. Reionisation is generally modelled as a step in the ionisation fraction, which must be a double step when taking the beginning and end of reionisation at different redshifts [207; 223], but the lower value of $\tau_{\rm R} \sim 0.05$ allows for a single step or a smooth transition from beginning to end. Looking at Table III–2.9 where $A_{\rm L}$ is allowed to vary, $\tau_{\rm R} = 0.023 \pm 0.012$ and $z_{\rm re} = 3.99 \pm 1.47$ are below the Gunn-Peterson bound.

The residual tension can again be quantified between the parameter value preferred by LSS compared to the CMB by introducing $\tau_{\rm R}^{\rm CMB,TT}$ and $\tau_{\rm R}^{\rm CMB,TE}$ in an MCMC fit. Here $\tau_{\rm R}^{\rm CMB,TT}$ is used for *Planck*2013 temperature data and $\tau_{\rm R}^{\rm CMB,TE}$ for large-angle WMAP polarization sourced at $z \lesssim 20$. The marginalised posterior of the difference, $\tau_{\rm R}^{\rm CMB,TE} - \tau_{\rm R}^{\rm CMB,TT}$, is non-zero at 2.1σ. Therefore, although *Planck*2013 temperature data is more compatible with $\tau_{\rm R}^{\rm CMB,TE}$ than

the active neutrino equivalent, one should also bear in mind the combined fit to CMB+LSS data of the varying $\tau_{\rm R}$ model is slightly worse. The likelihood ratio test shows the simpler one $\tau_{\rm R}$ model is preferred with a probability of 0.40%, again indicating the tension between $T_{\rm R}^{\rm CMB,TT}$ and $T_{\rm R}^{\rm CMB,TE}$ is real and not an artefact of overfitting.

In [4] the *Planck*2015 temperature anisotropies are combined with the `lollipop` likelihoods and obtains a lower value of the optical depth to reionisation $\tau_{\rm R}$. This shifts the *Planck*2015 value of $\tau_{\rm R} = 0.078 \pm 0.019$ to $\tau_{\rm R} = 0.058 \pm 0.012$ which is much closer to the $\tau_{\rm R} = 0.049 \pm 0.021$ from *Planck*2013+*All LSS (2013)* presented here from the 2013 study.

The discrepancy between the values of $\tau_{\rm R}$ inferred from WP and LSS suggests a possible resolution any source of tension. With the new, lower constraint on $\tau_{\rm R}$ from *Planck*+`lollipop` [4], the tension would be expected to reduce. Since the *Planck*+`lollipop` chains and likelihood code were not publicly available at present the *Planck*2015+Pol+BAO chains have been importance sampled using $\tau_{\rm R} = 0.058 \pm 0.012$. In this case, the quantification of tension when comparing to *All LSS (Weak)* reduces from $C = 0.550$ (0.76σ) to $C = 0.432$ (0.57σ). This also reduces to a minor extent from $C = 0.989$ (2.55σ) to $C = 0.985$ (2.44σ) for the comparison to *All LSS (Strong)*. At this stage it is not possible to make any conclusive statement that the lowering of $\tau_{\rm R}$ is in any more or less tension than *Planck*2015+Pol+BAO.

Chapter 3

Discussion

In this Part, several methods with which to quantify the amount of disagreement between parameter constraints when using two different data sets and the same model have been considered. Particularly, it has been seen that without knowing the precise details of the method being used, significantly different interpretations of *tension* can be made. In this Thesis, two new methods of quantification have been used which are robust in a wide variety of scenarios. These are the difference vector (*3*) and the integration between intervals (*4*) from Part III–1. Although they are not foolproof, a quantification by either of these methods should resemble what one would expect by comparing likelihood contours. More than this, they also work well in high dimensions and the interpretations remain easily understandable. This is unlike other methods, where consistent data sets could be interpreted as being significantly different or the results being very difficult to map to an easily understandable quantity describing the similarity of the distributions.

Armed with the new methods of quantification, a range of LSS probes were considered to constrain cosmological parameters for comparison with constraints from CMB observations. For independent LSS probes, joint constraints could be found which can be in either more *or* less tension than each of the individual constraints, depending on choice of analysis. From work presented from the 2013 analysis, when tight Gaussian priors were applied during MCMC analysis, the tension between the parameter constraints obtained when using the CMB and LSS was extremely significant. In the updated 2015 work presented in this Thesis, the priors were not applied and the tension was much less significant.

Assuming (whether correctly or not) the application of priors gave the correct discrepancy between LSS and CMB parameter constraints allowed the consid-

eration of extensions to ΛCDM. This included the addition of active or sterile neutrinos, which could be seen to reduce the tension, but not alleviate it completely. In the updated analysis, where the discordance was far reduced from the original work without any extensions, the addition of neutrinos did not affect the residual parameter distributions. Ad-hoc modifications to the primordial power spectrum were able to relieve the tension in the original work better than neutrinos, but this may have been due to the model's ability to fit spurious features in the CMB data, and still the tension was not wholly removed. The *Planck* lensing parameter was also allowed to vary to provide more correlated parameter space with neutrinos to be explored, finding that $A_{\rm L}$ has a preference for being larger than $A_{\rm L} = 1$ and that the active neutrino constraints were more significant. Finally, the effect of the WMAP polarisation data on the parameter constraints was considered. A significantly lower $\tau_{\rm R}$ is preferred when WP was ignored. This predicts the (now confirmed) results from the updated 2016 *Planck*+`lollipop` analysis [19].

Confirming that a discordance exists between data sets for a given model is difficult. Robust methods to quantify this needs considerable development. Indeed, if any tension does persist, a great deal of work is opened for consideration, whether this be improving the assumptions from given surveys or whether it be extending or changing the underlying cosmological model to solve the issue. This work on improving the robustness of comparative tests is essential for modern cosmology, perhaps taking precedence over the underlying physics which may be causing any deviations. As an example, it is more important that one understands whether differences in parameter values in ΛCDM are due to statistical interpretation than it is to understand whether neutrinos can be used to alleviate apparent discordances. To be able to introduce new physics it is imperative to understand how and why different statistical measures give different interpretations and gain a consensus comprehension of the results. These are all general topics which need to be considered in the future.

One particularly interesting subject, which is currently blossoming, is Bayesian hierarchical modelling [167]. Here, every single assumption made throughout the entire data processing, modelling and analysis has sets of *hyper-priors* describing the amount of belief in each uncertainty. This is extremely powerful since all unknowns are accounted for, highlighting any areas in the process where tensions in parameter distributions could come from. With this technique, it will be possible to understand whether current tensions have arisen from the incorrect assumptions made during analysis of the data, or perhaps more interestingly,

the presence of new physics which affects scales differently.

Appendix B

Comparison of methods

To understand how each of the different methods of quantification work it is useful to compare some simple distributions, shown in tables III–B.1 and III–B.2. Figures for each measure of every 1D and 2D parameter distribution comparison can be found on pages 139-141. As in Part III–1, these posterior distributions are $P_1 \equiv P(\theta|D_1, \mathcal{M})$ and $P_2 \equiv P(\theta|D_2, \mathcal{M})$ for data sets D_1 and D_2 respectively in a model $\mathcal{M}$. In the 1D case $\theta \equiv \theta$ is a one dimensional parameter, whereas in 2D $\theta = \{\theta_1, \theta_2\}$. In each case the probability distributions are normalised such that

$$\int d\theta P_i \quad = \quad 1. \tag{III–B.1}$$

	P_1	P_2
I	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$
II	$\mathcal{N}(0,1)$	$\mathcal{N}(0,3)$
III	$\mathcal{N}(5,1)$	$\mathcal{N}(-5,1)$
IV	$\mathcal{N}(0,1)$	$\mathcal{N}(1.427,1)$
V	$\mathcal{N}(0,1)+\mathcal{N}(-2,1)$	$\mathcal{N}(1.427,1)+\mathcal{N}(4,2)$

Table III–B.1: 1D probability distributions being compared.

I - Identical distributions Figure III–B.1 shows the distributions and integrated measures quantifying the amount of agreement or disagreement of two identical distributions, described in row I of tables III–B.1 and III–B.2. Each method is unanimous in its quantification of the combination of these two distributions in both 1D and 2D.

1, 2. The Bhattacharyya distance and the overlap coefficient are $B = 1$ and

	P_1	P_2
I	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$
II	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}3^2 & 0\\ 0 & 3^2\end{pmatrix}\right)$
III	$\mathcal{N}\left(\begin{pmatrix}5 & 5\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$	$\mathcal{N}\left(\begin{pmatrix}-5 & -5\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$
IV	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$	$\mathcal{N}\left(\begin{pmatrix}1.427 & 1.427\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$
V	$\mathcal{N}\left(\begin{pmatrix}0 & 0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$ $+\mathcal{N}\left(\begin{pmatrix}-2 & -2\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$	$\mathcal{N}\left(\begin{pmatrix}1.427 & 1.427\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)$ $+\mathcal{N}\left(\begin{pmatrix}4 & 4\end{pmatrix}, \begin{pmatrix}2^2 & 0\\ 0 & 2^2\end{pmatrix}\right)$

Table III–B.2: 2D probability distributions being compared.

$O = 1$, in both one and two dimensions. This shows the distributions are identical, since $P_1 = P_2$ then $\sqrt{P_1 P_2} = P_1 = P_2$ and $\text{Min}[P_1, P_2] = P_1 = P_2$ which is unity when integrated as in equations (III–1.1) and (III–1.2).

3. A value of $C = 0$ means that the distributions must be identical. The parameter ranges are identical identical distributions (and infinite for the distributions in tables III–B.1 and III–B.2) so the difference in the range is the same, $\delta\theta = \theta_1 = \theta_2$. A new Gaussian is formed with half the variance and a mean at $\delta\theta = 0$. Since $\delta\theta = 0$ is at the maximum of the distribution then there are no parameter ranges above the value of the probability distribution function at $\delta = 0$ to integrate. For the result in figure III–B.1 the result obtained by integrating inside the isocontour formed by the value of the probability density function at $\delta\theta = 0$ deviates slightly from zero due to the finite number of samples taken. The rest of the samples outside of this boundary can be considered consistent.

4. When $I_1 = I_2 = 0.997$ then the two distributions are shown to be identical. The set of parameter values which contain 99.7% of the samples of either distribution are equal for identical distributions. This means integrating either distribution for these parameter ranges will equate to $I_i = 0.997$, i.e. the total fraction of samples that can be drawn from the parameter ranges are drawn from both distributions.

5. There is no gain in information when two distributions are identical. Since this is expected it means there is also no *surprise*. This can be seen trivially in

equation (III–1.9) since $\log(P_2/P_1) = \log 1 = 0$ when $P_1 = P_2$.

6. A value of $\log R = 1.730$ or $\log R = 3.460$ shows that evidence favours the combined probability distribution when the distribution is chosen to be uniform between $-10 < \theta < 10$ in one or two dimensions. This is expected for identical distributions since, although the integrals $p(D_1)$ and $p(D_2)$ are greater than $p(D_1, D_2)$ their combination $p(D_1)p(D_2)$ is smaller. This is always true independent of the choice of prior. The magnitude of $\log R$ does depend on the prior: $\log R$ is larger when the prior is wider; it is smaller when the prior is narrower. The positive $\log R$ can be interpreted as an indication that the two distributions are somewhat similar. Although $\log R = 1.730$ *means* the distributions are identical with the given prior, it is not a particularly intuitive value.

7. Similar to measure *3* $\log T = 0$ shows that the two distributions are identical since $p(D_1, D_2)_{\text{shifted}} = p(D_1, D_2)$. Both of the means of the joint probability distributions are the same so the mean of the shifted distribution does not move. The ratio is therefore $T = 1$ giving a $\log T = 0$ showing that they are identical. This is again true in both 1D and 2D.

II - One distribution broader than the other but with the same mean

Figure III–B.2 shows the measure of discordance when one distribution remains the same as in I, but the width of the second distribution increases to $\sigma = 3$ as in the second row of tables III–B.1 and III–B.2. A useful measure here would indicate either that the distributions are very similar, or that P_1 is completely consistent with P_2 even though P_2 is not completely consistent with P_1.

1. $B = 0.775$ and $B = 0.600$ in one and two dimensions. These values show that the distributions are not concordant in some way. It does not illuminate in which way the distributions disagree. Knowing the distributions, it can be seen that the disagreement occurs because the value of P_2 are small for parameter values where P_1 is large, and vice versa. The integral over the combined distributions is therefore less than unity.

2. Similarly, the overlap coefficient reveals $O = 0.516$ and $O = 0.325$ in one and two dimensions respectively. The low maximum value of P_2 means that $\text{Min}[P_1, P_2]$ is capped where P_1 is large. This gives the same misleading interpretation as the Bhattacharyya distance. In fact, since the values of O are lower, they could be interpreted as the distributions being in greater disagreement.

3. The measure here does not take into account broadening of distributions and so $C = 0$ again. The variance of P_2 has increased (compared to in I) so the variance of the new distribution $P(\delta\theta)$ is larger, but the mean is still centred on $\delta\theta = 0$. The isocontour defined by the value of $P(0)$ contains no parameter values and so integrating again gives zero. This measure indicates that the samples

in the new distribution are *consistent* and so the original distributions agree. In fact, they can be interpreted as being identical, which may be misleading.

4. This measure is the most informative of all the quantifications of the level of agreement. $I_1 = 1.000$ and $I_2 = 0.684$ show that all of samples drawn from P_1 are contained in the parameter ranges which contain 99.7% of the samples drawn from P_2. Simply, P_1 is completely consistent with P_2. $I_2 < I_1$ indicates that P_2 has a greater variance than P_1, the value of I_2 showing how broad the distribution is in comparison to P_1. If $I_2 \lesssim I_1$ then P_2 is quite similar to P_1, but if $I_2 \ll I_1$ then P_2 has a much greater variance.

5. There is a gain in information from updating P_1 with P_2 since there is an extension of available parameter space, but this is mostly due to *surprise* as the entropy expected by broadening the distribution is small. On the other hand, when P_1 updates P_2 there is a much smaller relative entropy, but there is expected to be a large amount, so the *surprise* is negative. These two values can be interpreted as showing that P_2 does not agree with P_1 as much as expected and that P_1 agrees with P_2 *more* than is expected.

6. The interpretation of this measure is exactly the same as for I. The distributions must be similar since $\log R$ is positive. The value is lower for the same reason that the Bhattacharyya distance is less but, because it is normalised by the evidences of each distribution, it is still informative. As such it is possible to tell that, for a given prior, P_1 is not the same as P_2, but they are still similar.

7. Similar to measure *3*, $\log T = 0$ shows the distributions are consistent (or identical in fact). The maximum value of the distribution $p(D_1, D_2)_{\text{shifted}}$ is less, but it is still equal to $p(D_1, D_2)$ and so the logarithm of their ratio vanishes.

III - Discordant distributions Figure III–B.3 shows examples of each of the measures when two distributions are greatly separated. This is the last of the distribution combinations in which all of the measures are in agreement, showing that the distributions are not similar.

1, 2, 4, 6. Since P_1 is negligible where $P_2 \neq 0$ then the integration of any combination of P_1 and P_2 will (approximately) vanish, which explains the values of $B = 0$ and $O = 0$. Similarly, if the integration ranges where 99.7% of the samples from one distribution would be drawn do not overlap with the non-negligible regions of the other distribution then $I_i \approx 0$. Since $p(D_1)$ and $p(D_2)$ are much greater than $P(D_1, D_2)$ (which almost vanishes) then $\log R$ is extremely negative, preferring either evidence to the joint evidence. All these measures show that P_1 is not at all similar to P_2.

3. The mean of the new distribution is far $\delta\theta = 0$ and the value of the distribution is negligible there. The parameter range within the contour formed

where $P(0) = 0$ contains the whole distribution and as such $C = 1$. This is only possible when the whole distribution is integrated, showing that none of the samples drawn from either of the original distributions would be consistent with the other.

5. There is a very large relative entropy since the distributions contain completely different areas of parameter space, so a large amount of information is gained. However, since the means are incompatible, this information is not expected so the whole of the relative entropy is driven by *surprise*. This shows that the distributions do not agree with each other.

7. When means of P_2 are shifted to coincide with the means of P_1, $p(D_1, D_2) \ll p(D_1, D_2)_{\text{shifted}}$ and so T is large. A large positive $\log T$ indicates that the distributions are severely discordant.

IV - Slightly shifted distribution Figure III–B.4 shows the row IV distributions from tables III–B.1 and III–B.2. The second distribution P_2, has the same variance as P_1 but the means of P_2 are shifted such that the value of B is the same as using the distributions in row II of tables III–B.1 and III–B.2.

1. As already described, $B = 0.775$ and $B = 0.601$ in one and two dimensions. These are the same values obtained when the variance of P_2 is three times that of P_1. This example shows how the Bhattacharyya distance allows broadening of distributions to be mapped to shifts in the mean. Due to this, it is harder to interpret the meaning of B without seeing at least a projection of the probability distribution. $0 < B < 1$ could arise from purely a flattening of a distribution, or a shift in the means, or a combination of both.

2. The overlap coefficient is similar to the Bhattacharyya distance, although a shift in the means of one distribution is more heavily penalised (a lower value of O found) than a broadening of the variance of that distribution. The same problem still persists, that there is no distinction between flattening of the distribution or shifts or combinations of them both.

3. $P(\delta\theta)$ is centred slightly away from $\delta\theta = 0$ because the means of P_1 and P_2 are not equal. The value of the probability distribution at $\delta\theta = 0$ forms a contour (or interval) which contains 69.7% and 61.1% of the samples drawn from the distribution in one and two dimensions. These percentages can be mapped to the proportion of samples drawn from a one dimensional Gaussian, comparing the intervals to a number of standard deviations. In this Thesis, 69.7% would map to a tension of $\sim 1.0\sigma$, which means the distributions are consistent.

4. Since $I_1 = I_2 = 0.942$ then both P_1 and P_2 must have the same variance, but $I_1 = I_2 < 0.997$ shows that not all the possible samples are contained within the integration interval. This indicates that the means of P_1 must not coincide

with the means of P_2. Since the result of I_1 and I_2 are close to 0.997, then the means are not well separated and hence the distributions are in reasonable agreement.

5. The information gain from updating either distribution with the other is equal showing that both distributions have the same variance. In one dimension this gain is mostly expected and so the *surprise* is small and the distributions can be considered compatible. In two dimensions there is a lot more expected relative entropy than information gained and so the *surprise* is highly negative. This means the distributions are more similar *than expected*. It is difficult to quantify what this means in terms of similarity of the two distributions.

6. $\log R = 1.221$ and $\log R = 2.442$ in one and two dimensions. These values indicate that the joint evidence is more likely than each of the individual evidences $p(D_1)$ and $p(D_2)$, and therefore the distributions are similar. Interestingly, these measures show that the shift in the means of one of the distributions is *more* consistent than each of the distributions having equal means, but the variance of one being larger (as in row II of tables III–B.1 and III–B.2).

7. $\log T = 0.509$ and $\log T = 1.018$ shows that the distributions are similar but not identical, in one and two dimensions. The shifted joint evidence is slightly larger than $p(D_1, D_2)$, but because the means of P_2 are close to the means of P_1 the ratio between $p(D_1, D_2)_{\text{shifted}}$ and $p(D_1, D_2)$ is only slightly greater than one.

V - Unusually shaped distributions Figure III–B.5 shows the values each of the measures give for unusual shaped distributions (constructed by combining Gaussians in this case) in tables III–B.1 and III–B.2.

1, 2. The Bhattacharyya distance is lower than the comparisons of P_1 and P_2 in rows II and IV from tables III–B.1 and III–B.2 suggesting that these distributions agree less than in those cases. The same is true for the overlap coefficient. In the one dimensional case, mapping $B = 0.487$ to a shift in the mean only is equivalent to moving the peak of a Gaussian distribution by $\theta = 2.4$ from the centre of the other distribution. Likewise, $O = 0.264$ obtained here is equivalent to shifting the peak of a Gaussian distribution to $\theta = 2.2$ compared to another Gaussian with the same variance centred at $\theta = 0$. Comparing the values of B and O to shifts in the mean is a useful way to interpret results from these methods, although it still does not take into account the flattening of the distributions.

3. The values of $C = 0.620$ and $C = 0.310$ suggest that P_1 and P_2 are extremely consistent, although not identical. Mapping to one dimensional Gaussian distributions, these are equivalent to tensions of 0.9σ and 0.4σ respectively. This

maybe quite misleading since (according to figure III–B.5) a lot of the distribution lies away from $\delta\theta = 0$, it is just the primary peak which is near to $\delta\theta = 0$. This means it is the only measure here to quantify these two distributions as more consistent than in row IV of tables III–B.1 and III–B.2.

4. 70.4% of samples drawn from P_1 are within the the 99.7% confidence intervals of P_2 and 59.3% of the samples drawn from P_2 are within the 99.7% confidence intervals of P_1 in one dimension. This shows that P_1 is more consistent with P_2 than the other way around. Since $I_1 > I_2$ then the effective variance of P_2 is larger than P_1's. Both the values of I_1 and I_2 being less than 0.997 suggests a shift so that the peaks of the distribution are not aligned. Of course, the distributions could both be peaked at the same parameter value but one of the distributions skewed which would give similar results.

5. The relative entropy is mostly *surprise* driven suggesting the distributions are not in a great level of agreement. P_2 is less consistent with P_1 than P_1 is with P_2 since the information gain and *surprise* are smaller when P_1 is used to update P_2.

6. The positive values of $\log R = 0.120$ and $\log R = 1.110$ show that the two probability distributions are consistent since the joint evidence is more likely than either of the evidences combined. The values of $\log R$ are closer to zero than any of the previous comparisons from tables III–B.1 and III–B.2 with the exception of row III suggesting that the agreement is less in this case.

7. The ratio of $p(D_1, D_2)_{\text{shifted}}$ to $p(D_1, D_2)$ is fairly large so $\log T$ shows that the agreement is less than for the other comparisons in tables III–B.1 and III–B.2 except row III. The value is much less than $\log T$ for row III and so it is clear that these distributions are not wholly discordant.

When comparing the one and two dimensional distributions it can be seen that the general trends are the same. It should be noted here that the 2D distributions are slightly more distinct than the 1D distributions are for each row in tables III–B.1 and III–B.2 so the measure values are expected to show less consistency. The integration between interval (*4*) and difference vector methods (*3*) have the same interpretation value independent of the number of dimensions. The other methods (*1*, *2*, *5*, *6* and *7*) give different values in different dimensions, which needs to be taken into account or corrected when analysing the measures.

Probability distribution comparison figures

Each figure in this section shows the comparison of two probability distributions for each method discussed in Part III–1. The top row of figures III–B.1, III–B.2, III–B.3, III–B.4 and III–B.5 show the comparisons of the distributions in rows I, II, III, IV and V in table III–B.1 respectively. Likewise, the bottom row of each figure shows the comparison between the distributions in rows I–V in table III–B.2. The columns show the Bhattacharyya distance (*1*), the overlap coefficient (*2*), the integral of P_1 between the limits containing 99.7% of P_2 and the integral of P_2 between the limits containing 99.7% of P_1 (*4*), the quantification of Bayesian evidence (*6*), the shifted probability distribution (*7*), *surprise* (*5*) and the difference vector (*3*) from left to right. For the first six columns the solid, blue and dashed, red lines indicate the distributions P_1 and P_2 respectively. In the top rows, the shaded grey area (bounded by a dotted, black line) shows the integrated quantity used to give the comparison measure. In the bottom rows, the integrated quantities are shaded with blue being close to zero, turning red for $\mathrm{Max}[P_1, P_2]$. In the top row of the sixth column the green shaded area (bounded by a dot-dashed, green line) indicates the integrated shifted quantity $P_1 {P_2}^{\mathrm{shifted}}$, whilst the grey shaded area (bounded by a dotted, black line) marks the integrated non-shifted quantity $P_1 P_2$, the ratio of which gives the measure. The seventh column shows the amount of relative entropy in the wider, darker bars and the amount of surprise in the slimmer, lighter bars. The upper, blue bars indicate the relative entropy and surprise when P_2 is used to update P_1 and the lower, red bars show the relative entropy and surprise when P_1 updates P_2. The final column shows the probability distribution of the difference vector with a solid purple line. The grey shaded area in the top row is the integrated quantity giving the measure. The integration bounds are the values of the probability distribution greater than its value at $\delta\theta = 0$.

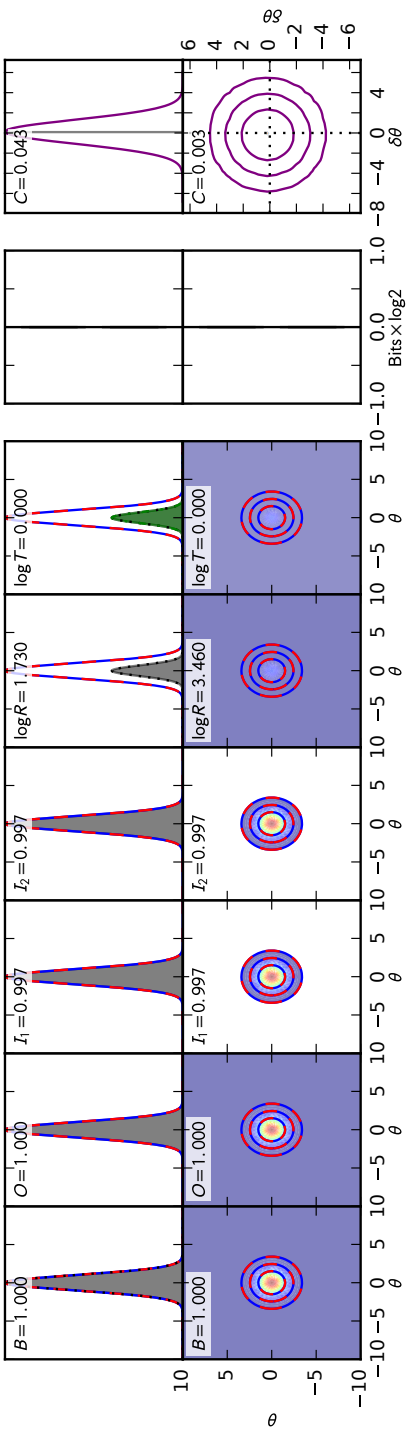


Figure III–B.1: Comparison of identical distributions (I)

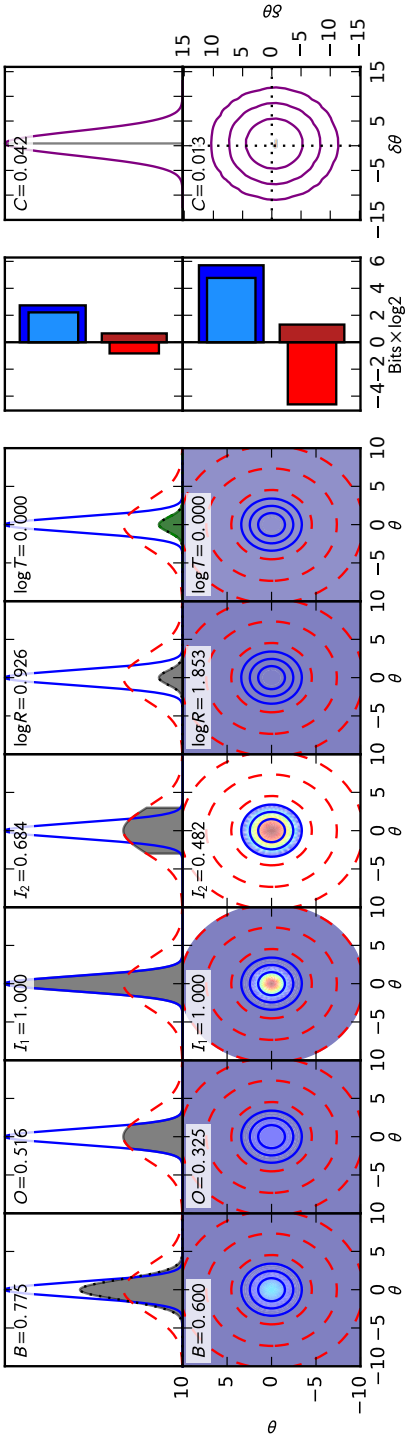


Figure III–B.2: Comparison of one distribution broader than the other but with the same means (II)

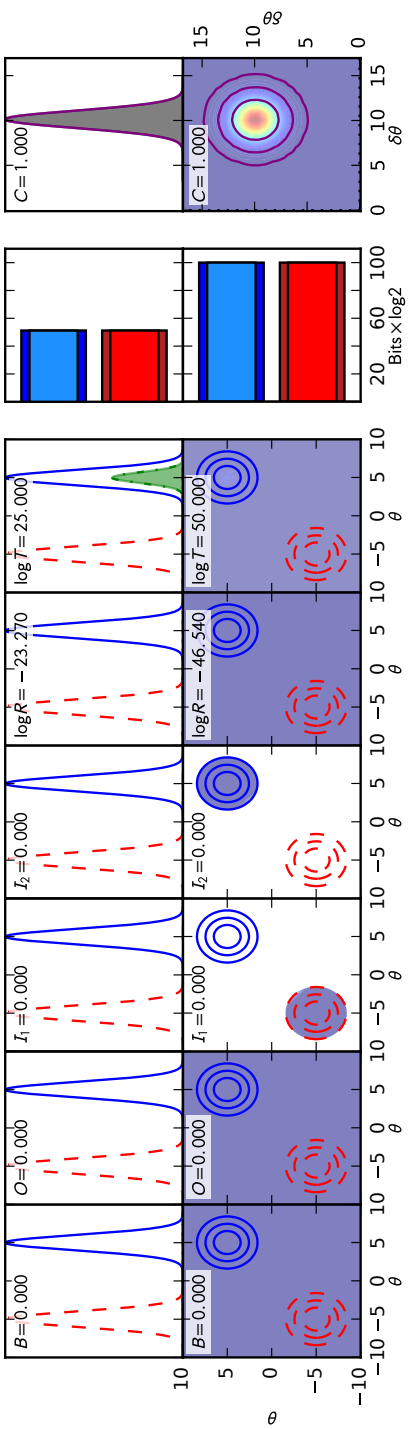


Figure III–B.3: Comparison of discordant probability distributions (III)

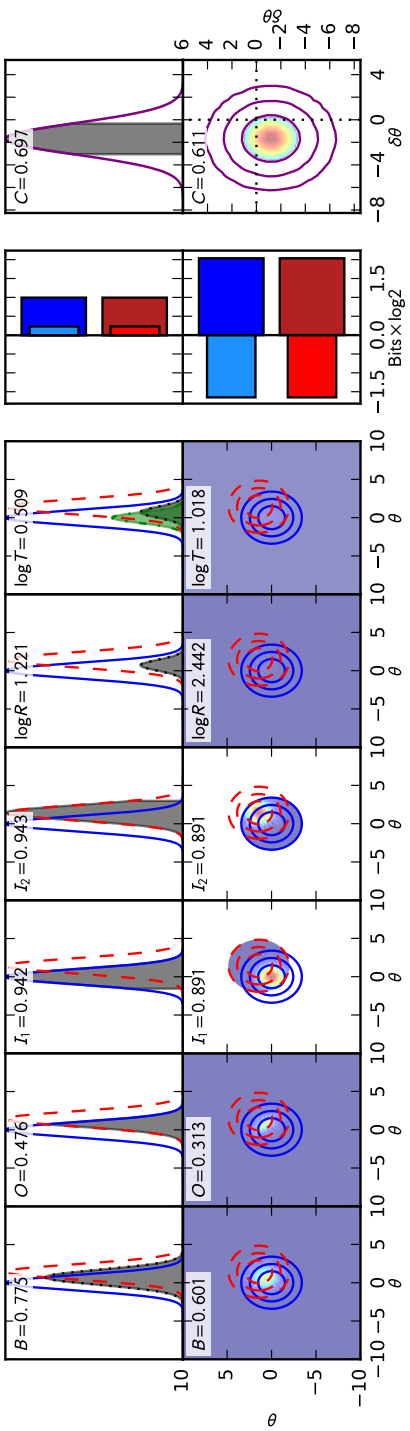


Figure III–B.4: Comparison of slightly shifted distributions (IV)

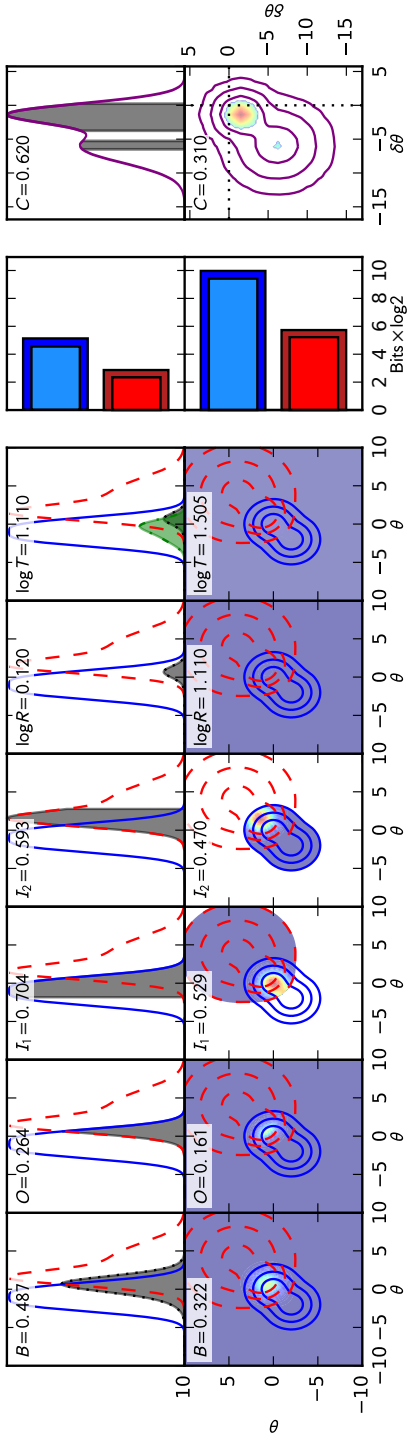


Figure III–B.5: Comparison of unusually shaped distributions (V)

Part IV

Deep learning

Deep learning is a powerful method for gaining multiple levels of abstraction, and has recently produced state-of-the-art results in tasks such as image classification and natural language processing [228]. It involves using machine learning to *learn* a specific algorithm, without needing to have any prior knowledge of what the algorithm should be like [321]. Machine learning itself is a study which dates back almost as far as the advent of computing [257; 286]. Based on early probabilistic ideas of the workings of biological neurons, starting with just a few generic functions, neural networks can be constructed to calculate solutions to problems which are almost intractable for computers. Computing power has, very recently, reached a point where large networks of neurons can be built, and along with advances in optimisation procedures, can now be used to solve many types of problems [134]. As well as image classification and natural language processing, neural networks are excellent at data mining [380] - searching for hidden correlations between inputs in large data sets - and autoencoding [70] - finding optimum compression procedures for any data type. Especially in the last few years, deep learning has been used by vast portions of the computer science community to attempt almost any task in computing where a large data set is available [52].

It is therefore interesting to consider the application of neural networks to the field of cosmology and astronomy. The field is very young, especially the use of deep learning. So far, a short list of possible attempts to solve astronomical problems includes: catalogue matching [305]; galaxy image classification [42; 341]; data belief networks [40]; data mining [39; 64]; and theoretical data prediction [134]. When coupled with the announcement and development of various large photometric surveys there are many applications of deep learning. These include, but are not limited, to the measurement of galaxy shapes from images, automated strong lens identification from multi-band images and galaxy cluster identification.

First, Part IV–1 contains a detailed description of machine learning, neural networks and how to train them. Then, in Part IV–2, a specific type of network is used to attempt to classify supernovae from the simulated raw data in the Supernovae Photometric Classification Challenge (SPCC) [213].

Chapter 1

Machine learning

1.1 Perceptrons

The most primitive artificial neurons are called perceptrons. They take a set of inputs x_i, and produce a single, binary output [257; 271; 286; 307]. Each of the inputs are connected to a node via a weight w^l_{ji}, where the j denotes the output node, i is the input node and l describes the layer (described in more detail below). Each node contains a bias b^l_i which is used to change the threshold at which the neuron *fires* [257]. It is useful to denote the weighted, biased input as

$$v^l_j = w^l_{ji} x_i + b^l_j \,, \tag{IV–1.1}$$

where there is implicit summing over repeated indices (i in this case). Whether or not the neuron is activated is dictated by a function ([286])

$$\phi(v^l_i) = \begin{cases} 0 & v^l_i \leq 0 \\ 1 & v^l_i > 0 \end{cases} . \tag{IV–1.2}$$

An example perceptron can be seen in figure IV–1.1 where $\phi(v_i)$ is taken to be the Heaviside function (almost) in equation (IV–1.2). As an example of the

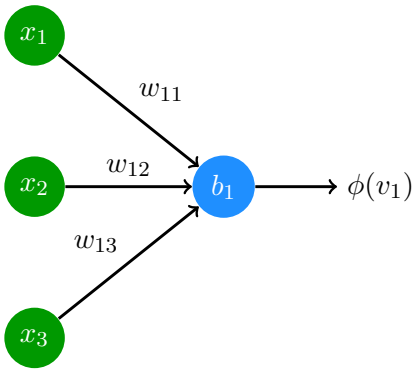


Figure IV–1.1: A simple, three input artificial neuron.

power of perceptrons, it is quite simple to construct a NAND gate from a network of just five neurons with two inputs each. Each of the ten weights are set to $w_{ij}^l = -2$ and the five biases are set to $b_i^l = 3$ [271]. The weights and biases do not need to be *trained*. Since a NAND gate can be built, any universal computation can be performed by combining series of these networks [97].

One problem with perceptrons is that a small change in the weights can cause a large change in the output, which makes it very difficult to *train* them to a specific algorithm [271]. The Heaviside-like function in equation (IV–1.2) can be modified to take values between 0 and 1. An *activation function* that can take on fractional values allows small changes in weights to propagate to small changes in the output. In fact, the change in the output $\Delta\phi(v_j^l)$, from node j is approximately a linear function of the change in all the weights Δw_{ji}^l from i to j and the change in the j node biases Δb_j^l ([271])

$$\Delta\phi(v_j^l) \approx \frac{\partial\phi(v_j^l)}{\partial w_{ji}^l}\Delta w_{ji}^l + \frac{\partial\phi(v_j^l)}{\partial b_j^l}\Delta b_j^l. \tag{IV–1.3}$$

It is quite straight-forward to calculate these using gradient descent, but first it is useful to discuss *hidden layers*.

1.2 Artificial neural networks

Deep learning depends on stacks (or layers) of neurons which are known as multilayer perceptrons (even when the activation function is non-binary) [97; 271]. The word *deep* implies a layer of abstraction between the inputs and outputs which cannot necessarily be understood [52; 147]. The inputs of the first layer become another set of inputs to a second layer, which are different from the original data, and so on through each of the hidden layers until the output. The output from each hidden layer is distinctly different (more abstract) than the original input layer [104]. It is useful to consider the input x_i as the output activation function of a *zeroth* layer $x_i = \phi(v_i^0)$. The outputs from each successive layer are then calculated as already prescribed such that the first hidden layer outputs are

$$\phi(v_j^1) = \phi(w_{ji}^1\phi(v_i^0) + b_j^1)\,, \tag{IV–1.4}$$

and the outputs from subsequent layers are ([271])

$$\phi(v_j^l) = \phi(w_{ji}^l\phi(v_i^{l-1}) + b_j^l)\,. \tag{IV–1.5}$$

This describes a feedforward artificial neural network (ANN), such as the one shown in figure IV–1.2. Other networks which do not simply feed the inputs

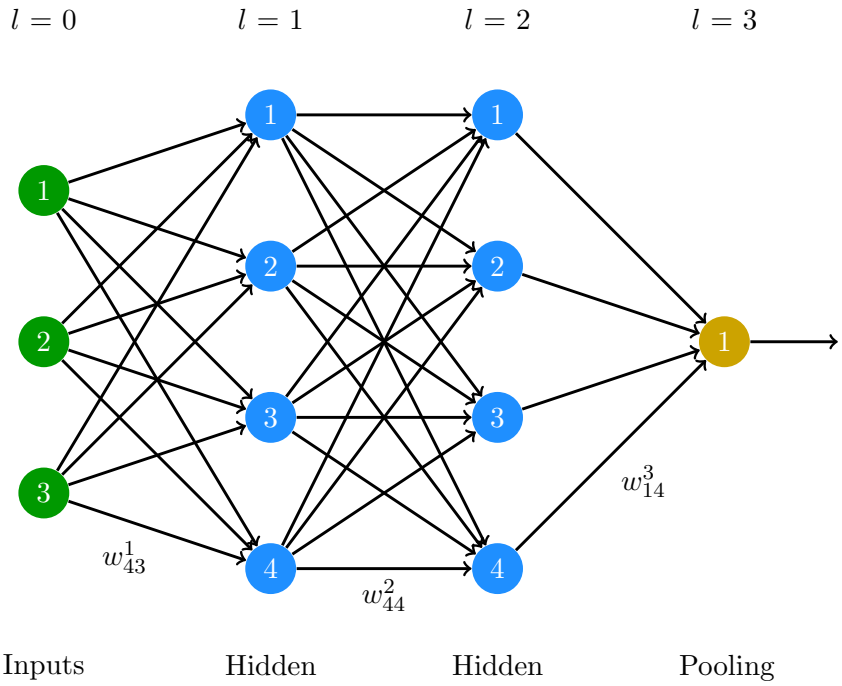


Figure IV–1.2: Full connected, two layer, ANN with four nodes in each layer, three inputs and one output.

forward through the network [128; 260] are possible. One example is a recurrent neural network (RNN), which is described in Part IV–1.5 and used for classification of supernovae in Part IV–2.

1.3 Back propagation

To approach the algorithm which the network is designed to recreate, the weights and the biases need to be set such that the inputs to the network result in the required output [271]. It is possible to train the network, which means the weights and biases do not need to be known *a priori* [311]. Training involves having a large set of training data with output results (in the supervised learning method described here). The input is fed through the network to the output and the weights and biases repeatedly nudged until the output from the network is representative of the pre-known output results (supplied as part of the data) [271]. As mentioned in equation (IV–1.3), the change in the output of each layer can be calculated as a linear function of the change in the weights and the change in the biases. This can be reversed, allowing the change in the weights and the biases to be calculated by knowing the change in the outputs for each layer. Since the output of the network needs to approach the correct result for the training data then a minimisation procedure can be performed on the error [311]. The

function which is going to be minimised needs to be sensitive to small changes in the output given small changes in the weights and biases. This *loss function* can take many forms, such as the mean square error between the data result and the network output, or the categorical cross-entropy [114]. For simplicity, consider the mean square error

$$L(\mathbf{w},\mathbf{b}) = \frac{1}{2n}\sum_{\mathbf{x}}^{n}||\mathbf{y}(\mathbf{x})-\phi(\mathbf{v}^L)||^2 \quad \text{(IV–1.6)}$$

where $\mathbf{w}$ and $\mathbf{b}$ are all the weights and biases in the network, $\mathbf{y}(\mathbf{x})$ is the vector of output results of each of the n vectors of training inputs to the network $\mathbf{x}$. $\phi(\mathbf{v}^L)$ is the vector of network outputs where L is the number of layers in the network.

It is useful to consider the change of the loss function given the weighted input at any layer v_j^l. Using the chain rule, relations can be made between the rate of change of the weights or the biases between different layers [311]. For the output layer the rate of change of the loss function with respect to the weighted, biased input is ([271])

$$\frac{\partial L}{\partial \mathbf{v}^L} = \frac{\partial L}{\partial \phi(\mathbf{v}^L)}\odot\frac{\partial \phi(\mathbf{v}^L)}{\partial \mathbf{v}^L} \quad \text{(IV–1.7)}$$

where $\odot$ is the Hadamard (element-wise) product [81]. $\partial L/\partial\phi(v_i^l)$ is small if neuron i is unimportant [271]. To propagate the rate of change of the loss function with respect to the weighted, biased input to any prior layer l

$$\frac{\partial L}{\partial \mathbf{v}^l} = \left(\mathbf{w}^{l+1}\right)^T\frac{\partial L}{\partial \mathbf{v}^{l+1}}\odot\frac{\partial \phi(\mathbf{v}^l)}{\partial \mathbf{v}^l}\,. \quad \text{(IV–1.8)}$$

$(\mathbf{w}^{l+1})^T$ is the transpose of the weights at the $l+1$ layer [271]. The rate of change of the loss function with respect to the weighted, biased input at the $l+1$ layer is propagated by the transpose of the weights to layer l. The Hadamard product propagates this error through the l layer activation function to give the rate of change of the weighted, biased output at layer l. Using these two equations, the rate of change of loss function with respect to the weighted, biased output at any layer can be found [271].

To optimally train the weights and biases in the network the change in the loss function must be negative, i.e. the loss function approaches its minimum [311]. Since the change in the loss function given a change in the weights or the biases is simply

$$\Delta L \approx \frac{\partial L}{\partial \mathbf{w}^l}\Delta\mathbf{w}^l\,, \quad \text{(IV–1.9)}$$

and

$$\Delta L \approx \frac{\partial L}{\partial \mathbf{b}^l}\Delta \mathbf{b}^l\,, \tag{IV–1.10}$$

then choosing

$$\Delta \mathbf{w}^l = -\eta\frac{\partial L}{\partial \mathbf{w}^l}\,, \tag{IV–1.11}$$

and

$$\Delta \mathbf{b}^l = -\eta\frac{\partial L}{\partial \mathbf{b}^l}\,, \tag{IV–1.12}$$

means that $\Delta L = -\eta||\partial L/\partial \mathbf{w}^l||^2 \leq 0$ and $\Delta L = -\eta||\partial L/\partial \mathbf{b}^l||^2 \leq 0$ respectively [271; 311]. η is a learning rate which must be small so that the approximation in equations (IV–1.9) and (IV–1.10) is respected, but must be large enough that the weights will update [271]. The transformation in the weights and biases which will optimally cause the loss function to reduce on every pass of training input are ([311])

$$\mathbf{w}^l \rightarrow \mathbf{w}^l - \eta\frac{\partial L}{\partial \mathbf{w}^l}\,, \tag{IV–1.13}$$

and

$$\mathbf{b}^l \rightarrow \mathbf{b}^l - \eta\frac{\partial L}{\partial \mathbf{b}^l}\,. \tag{IV–1.14}$$

These equations can be used to define the transformations of the weights and biases to achieve a change in the output. Again using the chain rule, the relation between the rate of change of the loss function with respect to the weights or with respect to the biases and the quantity found in equation (IV–1.8) can be calculated

$$\frac{\partial L}{\partial \mathbf{v}^l} = \frac{1}{\phi(\mathbf{v}^{l-1})}\frac{\partial L}{\partial \mathbf{w}^l}\,, \tag{IV–1.15}$$

$$\frac{\partial L}{\partial \mathbf{v}^l} = \frac{\partial L}{\partial \mathbf{b}^l}\,. \tag{IV–1.16}$$

Back propagation is an extremely efficient method of updating the weights and biases since it is equivalent to only two passes through the network (with some overhead from calculating the derivative of the activation function) [271]. The full training algorithm for a single piece of training data is

1. For every piece of data in the training set input the vector $\mathbf{x} = \phi(\mathbf{v}^0)$, into the network.

2. Feed forward and compute the weighted, biased input in equation (IV–1.1) and the activated output from equation (IV–1.5) at every layer for every training set input.

3. Compute the rate of change of the loss function (which is averaged over the whole training set) for the weighted, biased output in the last layer using equation (IV–1.7).

4. Back propagate through the entire network to find $\partial L/\partial \mathbf{v}^l$ at every layer using equation (IV–1.8).

5. Calculate the gradients of the loss function with respect to the weights and the biases respectively using equations (IV–1.15) and (IV–1.16).

6. Update the weights and biases using equations (IV–1.13) and (IV–1.14).

In practice, carrying out this process is quite computationally heavy due to using the whole training data on every update of the weights. A technique called stochastic gradient descent can be used to speed up the training [139]. Here, the training set is randomly split into *batches* (often called mini-batches) of m sets of inputs. A good estimate of the gradient of the loss function with respect to the weighted, biased inputs can be calculated from the mini-batch. This can then be used to update the weights and biases and repeated for all batches in the training set [271]. This completes one *epoch* of training at which point the training set can be randomly split into a different set of batches and training commence as before.

If the weights are all initialised to the same value (say zero) before training then there is no asymmetry in the network which can make large portions of the network redundant [373]. For example, since the output of each neuron will be the same then back propagation will update the final layer of a fully connected ANN (such as in figure IV–1.2) to the same weights. Back propagation will again update the previous hidden layers to similar weight values all the way back to the input. Only after the first update to the weights and biases will the inputs start having an effect on the output. By initialising each set of weights to a very small value from a Gaussian distribution about zero then there will be some random asymmetry in the network which can encourage more constructive back propagation of the loss function [373]. Since the variance of the outputs grow with greater number of inputs, the variance of the Gaussian to draw weights from should be normalised to $\sigma = 2/n$ where n is the number of inputs [373].

1.4 Activation functions

The activation functions can take on many forms, a few of the popular ones are the sigmoid, hyperbolic tangent, rectified linear unit (ReLU) and `softmax`

functions: $(1+\exp[-v_i])^{-1}$, $\tanh(v_i)$, $\text{Max}[0, v_i]$ and $\ln(1+\exp[v_i])$ [97; 166; 225; 271]. Examples of these are shown in figure IV–1.3.

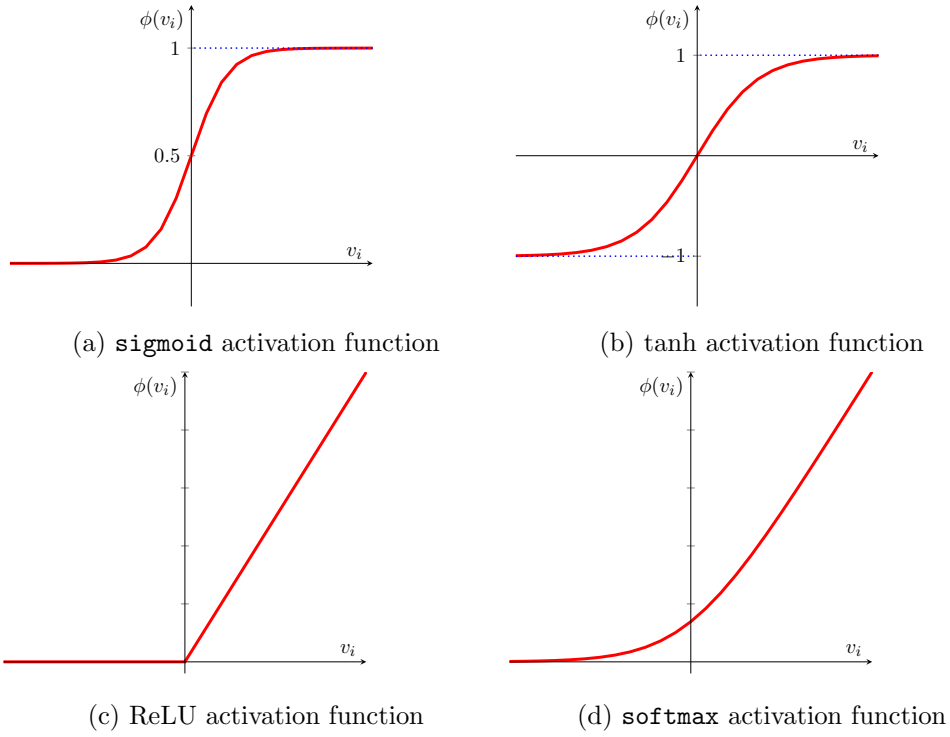
Although the sigmoid function most closely matches the activation of biological neurons, it is widely regarded as dangerous for training purposes since the gradient vanishes when $\phi(v_i^l) \approx 0$ or $\phi(v_i^l) \approx 1$ [306]. The neuron *saturates* and stops learning, i.e. $\partial L/\partial w_{ji}^l \approx 0$ so equation (IV–1.13) means $w_{ji}^l \to w_{ji}^l$. The same argument is true for the biases. The sigmoid function also is non-zero centred, which can affect the gradient descent during training. The hyperbolic tangent activation function is therefore preferred over the sigmoid [229]. Although it saturates, it is zero-centred which mitigates some of the training issues. Both of these functions were originally thought to be the most useful for mapping outputs to probabilities [271].

Taking the Max value of the weighted input has many advantages over the sigmoid or hyperbolic tangent function [267]. Training using this activation function can be up to six times faster than the other two methods [225]. This is both due to the speed of the calculation of the Max function as well as its ease of use in optimisation procedures. Unfortunately ReLU has a tendency to set weights to zero so that the activation function never fires for a large proportion of the neurons, which means the network becomes extremely inefficient and cannot be trained beyond a certain point [166]. This is because the gradient of the function when $v_i^l < 0$ is zero and hence the weights and the biases stop getting updated. The `softmax` function partially improves on the common ReLU since it is a smooth function and can be analytically differentiated (to the sigmoid function). For some architectures this is quicker to calculate than the gradient of the Max function, but not always [349]. There is still the issue of large portions of the activation functions not firing after training since the gradient vanishes for more negative values of v_i^l.

More recent developments in activation functions include the leaky ReLU [244] and parameterised ReLU (PReLU) [166]. The former activation function has a pre-defined small negative slope

$$\phi(v_i) = \begin{cases} \alpha v_i & v_i < 0 \\ v_i & v_i \geq 0 \end{cases} . \tag{IV–1.17}$$

This prevents the gradient from vanishing, but the value of α needs to be picked correctly to ensure optimal training. There is no intuitive way to achieve this. Instead, α can be included as a tuneable parameter along with the network weights. This gives rise to PReLU activation functions [166]. Each neuron in the



(a) `sigmoid` activation function

(b) tanh activation function

(c) ReLU activation function

(d) `softmax` activation function

Figure IV–1.3: Example activation functions

network has a different value of α^l (like an additional bias) so that each neuron is independent of the input data used during training. This means that there is much less chance of the neuron saturating. Finally, randomised ReLU (RReLU) activation functions work by choosing a random value of α for every batch of training, and is then set to a constant value during training [383]. Although this may seem counterintuitive, the freedom in the network not knowing the final value of α prevents overfitting (described below) and so performs better than the previous methods. The use of the ReLU family of activation functions is much preferred over either the sigmoid or hyperbolic tangent function [271].

1.5 Recurrent neural networks

Knowing about ANN, back propagation and different types of activation functions allows more interesting networks to be considered. As already briefly mentioned in Part IV–1.2, RNN are one such interesting application of neural networks, which is not simply feed forward [128; 238; 279; 320]. Here there are feedback loops between neurons which do not affect the training given one set of inputs, but can cause the neurons to fire differently on the next set of inputs, i.e. causing the network to learn about sequences in sets of data as well as about the inputs themselves. This a very important feature for extended data sequences,

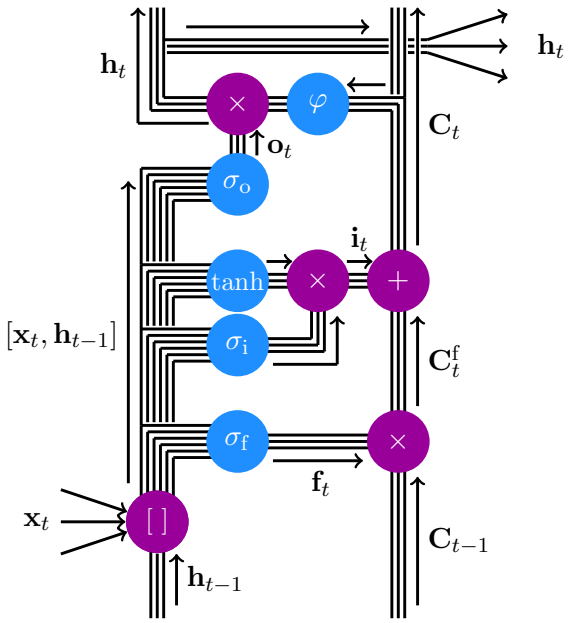


Figure IV–1.4: An LSTM cell showing the functions which allow the state of the network to be remembered over long sequences.

say natural language processing. It is difficult to process the meaning of a sentence by learning each word in the sentence separately (even for humans). On the other hand, an understanding of the sentence can be achieved by looking at the sentence as a whole. To achieve this, the influence of previous words in the sentence need to be tracked. It is long term correlations, like the sequence of words in sentences, that RNN architecture can learn about.

In general, RNN can only learn about data close in the sequence by passing the activated output of the network back as an additional *hidden* input for the next input in the sequence [260]. This short term knowledge can be improved upon using a specific neuron type called a long short term memory (LSTM) unit [179], or variants like the gated recurrent unit (GRU) [84], described below and used in Part IV–2.

An LSTM cell is a particular set of functions which is capable of storing important information from the sequence of the input vectors over extended regions of the sequence [272].

A hidden state $\mathbf{h}_{t-1}$ contains the LSTM cell output vector from the previous step in the sequence $t-1$. This is combined with the input vector from the network at sequence step t, $\mathbf{x}_t$ [179]. This combination can be performed in several ways, two popular forms are concatenation and consensus. When using

concatenation $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$ pass through two identical copies of the network and so the only influence the prior state has on the input is the shared weights of the functions in the cell [260]. Consensus combination simply adds the values (or in some cases finds the average) of $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$. Figure IV–1.4 shows an example LSTM cell where the input and hidden state are concatenated to form $[\mathbf{x}_t, \mathbf{h}_{t-1}]$.

First, $[\mathbf{x}_t, \mathbf{h}_{t-1}]$ is passed through a *forget gate* [141] which is a sigmoid function. Note this sigmoid function is not a neuron, since there are as many outputs and inputs

$$\mathbf{f}_t = \sigma_{\mathrm{f}}(\mathbf{w}^{\mathrm{f}} \odot \mathbf{x}_t + \mathbf{u}^{\mathrm{f}} \odot \mathbf{h}_{t-1} + \mathbf{b}^{\mathrm{f}}). \tag{IV–1.18}$$

Each element in the input vector $x_{i,t}$ has a weight w_i^{f} and each element in the hidden state vector $h_{i,t-1}$ has a weight u_i^{f} [272]. The sigmoid function also has a bias for each element in either of the vectors b_i^{f}. The output of this sigmoid function contains information about what is important and not in both the hidden state and the network input (represented by $\mathbf{f}_t = 1$ and $\mathbf{f}_t = 0$ respectively and values in between) [141]. This *forget* vector is multiplied by the state vector from the previous sequence step $\mathbf{C}_{t-1}$, which contains all the important information about each element in the input vector over the sequence up to step $t-1$,

$$\mathbf{C}_t^{\mathrm{f}} = \mathbf{f}_t \odot \mathbf{C}_{t-1}. \tag{IV–1.19}$$

The multiplication removes (or forgets) unimportant information in the state vector and weights down less important features [141; 272].

$[\mathbf{x}_t, \mathbf{h}_{t-1}]$ is also passed through another sigmoid gate to select for importance and a hyperbolic tangent function separately, and combined via multiplication to create the vector of important inputs at sequence t

$$\begin{aligned} \mathbf{i}_t = {} & \sigma_{\mathrm{i}}(\mathbf{w}^{\sigma_{\mathrm{i}}} \odot \mathbf{x}_t + \mathbf{u}^{\sigma_{\mathrm{i}}} \odot \mathbf{h}_{t-1} + \mathbf{b}^{\sigma_{\mathrm{i}}}) \\ & \odot \ \tanh(\mathbf{w}^{\mathrm{i}} \odot \mathbf{x}_t + \mathbf{u}^{\mathrm{i}} \odot \mathbf{h}_{t-1} + \mathbf{b}^{\mathrm{i}}), \end{aligned} \tag{IV–1.20}$$

with another four vectors of weights $w_i^{\sigma_{\mathrm{i}}}$, w_i^{i}, $u_i^{\sigma_{\mathrm{i}}}$ and u_i^{i}, and biases $b_i^{\sigma_{\mathrm{i}}}$ and b_i^{i} for each element in $[\mathbf{x}_t, \mathbf{h}_{t-1}]$ [179]. This is then added to the state vector which has forgotten unimportant information, creating

$$\mathbf{C}_t = \mathbf{C}_t^{\mathrm{f}} + \mathbf{i}_t. \tag{IV–1.21}$$

$\mathbf{C}_t$ is the state vector which contains all the important information from all sequence steps including the current step t [179].

Finally an output sigmoid is used to select important features, with a new set of weights w_i^{o}, u_i^{o} and biases b_i^{o} ([179; 272])

$$\mathbf{o}_t = \sigma_{\mathrm{o}}(\mathbf{w}^{\mathrm{o}} \odot \mathbf{x}_t + \mathbf{u}^{\mathrm{o}} \odot \mathbf{h}_{t-1} + \mathbf{b}^{\mathrm{o}}). \tag{IV–1.22}$$

$\mathbf{C}_t$ is passed through an activation-like function $\varphi(\mathbf{C}_t)$ to force the values in the vector to be of the form needed for the output [179]. φ can be any activation-like function, say the hyperbolic tangent function or a ReLU, but with an output for every input [272]. The output vector is then given by

$$\mathbf{h}_t = \mathbf{o}_t \odot \varphi(\mathbf{C}_t), \tag{IV–1.23}$$

creating the current hidden state vector which contains only the important information from the current sequence step [179; 272]. Two copies of $\mathbf{h}_t$ are made, one is passed back to the LSTM cell as $\mathbf{h}_{t-1}$ on the next step, and the other is outputted from the cell. It is usual that this output goes through an activation (in the neuron sense) so that the LSTM cell also acts as a neuron, with one output [83]. Most current uses of LSTM cells have ReLU activation on the output.

Since there are essentially three inputs from the network (passing through functions $\mathbf{f}_t$, $\mathbf{i}_t$ and $\mathbf{o}_t$) then back propagation can set the weight of each gate differently. This allows the sequence to be learned about, i.e. which inputs to forget, which inputs to propagate through the sequence and which inputs to pass through the network at a given sequence step [141; 179].

GRU are a more recent development of the traditional LSTM, where the forget and importance gates are combined and the hidden state vector and cell state are not distinguished between [84]. This means there are fewer weights to set, and so training is quicker, but GRU can have less freedom than traditional LSTM. More complicated network architecture tends to be needed when using GRU to overcome the reduced number of weights.

1.6 Data augmentation

Since extremely large amounts of data are needed for training a network, especially deep networks, the input data can be augmented to increase the training set size. Data augmentation takes one of the original pieces of data and modifies it in some way and uses both the original and modified data as input to the network [380]. For example, if using an image as an input, it could be augmented by flipping the image across either the horizontal or vertical axis, rotating the image, adding a small amount of noise to the pixel values, randomly cropping

etc [373]. By doing this, a similar image is produced which the network can learn about, without over-learning about the specifics in the original image.

1.7 Overfitting

Overfitting occurs when a network learns about specific features in the training set, which may not be present in the test set [113]. In particular, sampling noise may be learned about, i.e. noise in the training set may statistically correlate but not be present in the entire data when the test set is included [164]. The weights in the network are set to detect these features, and so sets preference for these features in the test set. The training therefore over-learns this noise and cannot learn more about the real features in the data. Overfitting can typically be detected by comparing the loss of the training and test data. If the loss of training data continues to decrease, but the loss of the test data increases, this is a sure sign of overfitting [355]. If, on the other hand, no sign of overfitting is observed, the network is not usually complex enough to fully learn the relationship between inputs and outputs (called *underfitting*) [355].

To prevent overfitting regularisers can be added [373]. The most common is the $L2$ regulariser where the loss function is appended with $\lambda w_{ji}^2/2$, for every weight in the network, where λ is a *strength* parameter for the regulariser [256]. This means that particularly strong weights are penalised and less strong weights preferred. Overfitting is therefore prevented, since the "well known" features (statistically correlated noise) in the training set will become less important than the overall (real) features of the set [373]. This is necessary because the well known features in the training set might not end up being in the test set and therefore the network will prefer knowledge which it doesn't need.

Another way to prevent overfitting is $L1$ regularisation, where the loss function has $\lambda|w_{ij}|$ added to it [256]. Almost all weights are set to zero, making a particularly sparse network which picks up only very specific features in the inputs regardless of noise [373]. This usually makes training the network more difficult than when using $L2$ regularisation since fewer features from the inputs can be considered by the network.

By constraining the weights such that they never exceed a value c, then the network can be prevented from over-learning one specific input, where the weight becomes extremely large [373]. Since the network weights cannot become too large, no single weight can dominate the network, although specific features can

be learned better than for the $L2$ regulariser. c is arbitrary and it can be hard to tune this parameter to best constrain the network.

Finally, dropout is an extremely useful technique which can be used by itself or as a complement to the other regularisation techniques [344]. During training a fraction of the network weights are set to zero and only the non-zero fraction of the weights are updated for each batch. This effectively samples different networks in the total number of possible networks for the given architecture [344]. Training a network with n neurons when using dropout is equivalent to training 2^n thinned networks (with weight sharing) to learn about different features. Each network may learn about specific features, but the ensemble does not know about these features and so overfitting can be prevented. Dropout is the technique used in Part IV–2.3 to prevent overfitting.

Chapter 2

Supernovae classification

In this chapter supernovae will be classified using deep RNN. LSST, for example, is expected to find over 10^7 supernova [2]. However, it is estimated that only 5000 to 10,000 (although these numbers are not guaranteed) will be spectroscopically confirmed by follow up surveys [255], so classification methods need to be developed for photometry. All previous approaches to automated classification [208; 242; 269] have first extracted features from supernovae light curves before using machine learning algorithms. One of the advantages of deep learning is replacing this feature extraction.

Supervised deep learning is used here. During training, the machine is given inputs and produces a set of output predictions. It is also given the correct set of outputs. An objective loss function then measures the error between the predicted and target outputs, and the machine updates its adjustable parameters to reduce the error. It can then make predictions for unknown outputs.
RNN are a class of ANN that can learn about sequential data [260]. They are commonly used for tasks such as speech recognition and language translation, but have several possible applications in astronomy and cosmology for processing temporal or spatial sequential data. RNN have several properties which makes them suitable for sequential information. The inputs to the network are flexible, and they are able to recognise patterns with noisy data (for example the context of a word in a sentence relative to others can vary, or a time stream can contain instrument noise).

As described in Part IV–1.5 the main problem with vanilla RNN is that they are unable to store long term information, so inputs at the end of a sequence have no knowledge of inputs at the start. This is a problem if the data has long term correlations. LSTM units can be used to overcome this problem. A detailed

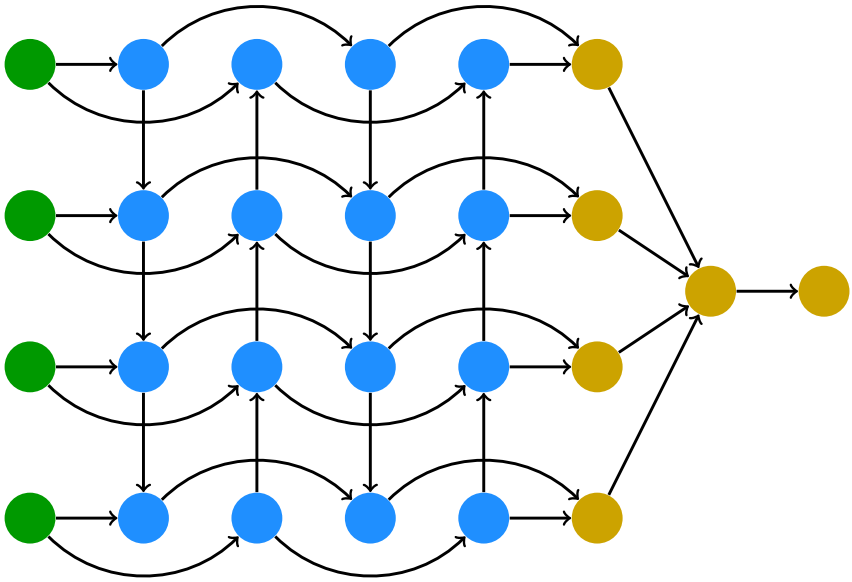


Figure IV–2.1: Two-layer, bidirectional RNN for sequence classification. The input vectors at each sequential step are fed into a pair of bidirectional hidden layers, which can propagate information forwards and backwards. These are then merged to obtain a consensus view of the network, and finally a `softmax` layer computes classification probabilities.

description of LSTM [179] is presented in Part IV–1.5. Another problem with RNN is that information can only flow in one direction. With *bidirectional* RNN, information is able to pass both forwards and backwards. Bidirectional LSTM networks have been shown to be particularly powerful where sequential data is accompanied by a set of discrete labels.

The architecture of a typical bidirectional RNN for sequence labelling is shown in figure IV–2.1, where the circles represent neurons. In this case the inputs, which are vectors at each sequential step, are connected to two hidden RNN layers, either vanilla RNN or memory units. Each hidden layer contains a number of hidden units (capable of storing information), and in each layer information flows either forwards or backwards, but no information passes between the two directions. Several hidden layers can be stacked to form *deep* neural networks. Deep networks are capable of learning higher-level temporal or spatial representations, and complex relationships between the inputs and outputs.

The output from the final set of hidden layers in each direction is merged at each sequential step, and mean pooled (averaged) over all steps to obtain a consensus view of the network to improve the performance of the network. Finally, the mean output is fed to a `softmax` layer, taking an input vector $\phi(\mathbf{v}^{L-1})$ and returning normalised, exponential outputs for each class label i, $\exp(\phi(v_i^{L-1}))/\sum_i \exp(\phi(v_i^{L-1}))$, i.e. a vector of probabilities.

Each neuron is connected to another by a weight matrix, and the optimal weights are found by back propagating the errors from a *loss function* of the output layer. For classification problems, this is typically the categorical cross-entropy between predictions and targets, defined as

$$L \quad = \quad -t_{i,j} \ln (p_{i,j}) \tag{IV–2.1}$$

where i, j run over the class labels, $t_{i,j}$ are the targets for each class (either 0 or 1) and $p_{i,j}$ are the predicted probabilities. Back propagation takes the derivative of the loss with respect to the weights $\mathbf{w}^L$ of the output layer, $\partial L/\partial \mathbf{w}^L$, the chain rule is used to update the weights in the network as described in Part IV–1.

2.1 Example data

Data from the SPCC [212; 213] is considered here, consisting of 21,319 simulated supernova light curves. Each supernovae sample consists of a time series of flux measurements, with errors, in the g, r, i, z bands (one band for each timestep), along with the position on the sky and dust extinction. An example set of light curves is shown in figure IV–2.2.

Due to the format of the input data, a small amount of data processing is first performed to obtain values of the g, r, i, z fluxes and errors at each sequential step. This assumes the time sequence begins at day 0 for each supernovae, rather than counting days forwards and backwards from the maxima of the light curve. For observations less than ~ 1 hour apart, the g, r, i, z values are grouped into a single vector, ensuring there is at most one filter-type in each group. If there is more than one filter-type, the group is further subdivided using a finer time interval. The group time is the mean of the times of each observation, which is reasonable as the time intervals are small compared to the characteristic time of the light curve.

Figure IV–2.3 shows how the length of the grouped-time data vector is related to the duration of the light curve. The bottom left subplot shows that when

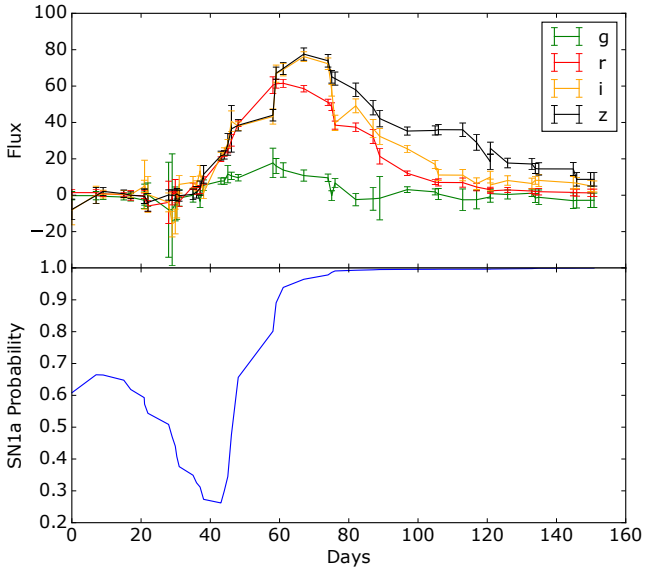


Figure IV–2.2: (Top) Example light curve in the four g, r, i, z bands for SN ID 551675 (a type-Ia) in the SPCC data [213]. The data has been processed using augmentation so there is a g, r, i, z value at each sequential step. (Bottom) Type-Ia probability as a function of time from a two-layer LSTM model, trained with around 10^4 supernovae and SN 551675 excluded. The final probability gives 99.5% confidence that the supernovae is of type-Ia.

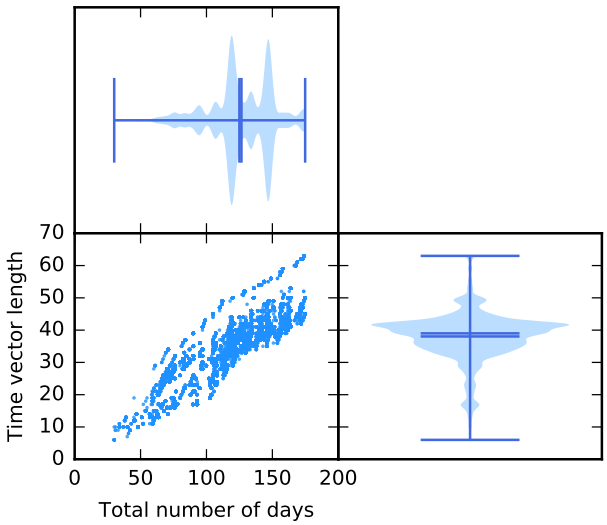


Figure IV–2.3: . (Top) Distribution of the total number of days for each light curve with the minimum, maximum, mean and median values indicated. (Bottom right) Distribution of the number of elements in the grouped time vector with the minimum, maximum, mean and median values indicated. (Bottom left) The trend showing that more days in the light curve result in longer group time vectors.

Time	g	r	i	z
t_1	g_1	r_1	i_1	z_1
t_2	g_2	r_2	$-$	z_2
t_3	g_3	r_3	i_3	z_3

Table IV–2.1: Data augmentation of missing observations. The missing data is replaced randomly by a value between i_1 and i_3.

there are more days since the beginning of observation of the light curve results there are a greater number of grouped time elements in the vector. The upper subplot shows the distribution of observation lengths in the SPCC data varies significantly with two distinct peaks. These are grouped into an average of 40-element data vectors as can be seen in the bottom right subplot.

Observations are of the form in table IV–2.1, where any missing values are denoted by a dash. In order to impute the missing value of i, *data augmentation* is used, by randomly selecting a value between i_1 and i_3. Five random augmentations of all missing data are made, thereby increasing the size of the data set fivefold. The importance of this can be tested by training each augmentation separately and comparing the change in accuracy, which is $\sim 1\%$. Training with multiple augmentations at once gives the best performance since the network learns to ignore random-filled values.

The data comes in two types, those with and those without the host galaxy photometric redshift. Each data set is split into a training and test set, with the training set containing a spectroscopically confirmed supernovae type and redshift. It is important that augmented data with the same supernovae ID go into either the training *or* test set otherwise they will not be independent. The original SPCC data consisted of 1,103 training samples. The answer keys were subsequently made available for the test set [212].

The input vector to each sequential step consists of: time in days since the first observation; flux in each of the 4 bands; flux errors in each of the 4 bands; RA and Dec; dust extinction; and host photo-z if relevant. Whilst some of these variables are not expected to impact the classifier accuracy, the network is left to decide if they are relevant, preventing any feature engineering.

RNN typically perform better with more training data, so the network is trained using the SPCC test set with answer keys (which is a non-biased representational

data set - unlike the original SPCC training set was non-representational), and a random fraction selected to act as the training set. 1,103 supernovae (a training fraction of 0.052) are considered, the same size as the original challenge, and fractions of 0.25 and 0.5 (around 5000 and 10^4 supernovae respectively), nearly an order of magnitude larger, and closer to the number likely to be followed up for LSST. The training performance of RNN is also improved if the data is processed in mini-batches. In order to do this the input data must be of the same length, so the sequence length is set to be the maximum length over all supernovae observations, the input is prepended with padding. In training the network the padding is ignored by masking the padded input.

The times of the observations in the light curve are irregularly spaced and whilst this may not be optimal for the network it is found that it is better to use the data padded at the end of the sequence than to place observations at similar times in similar sequence positions. There may even be hidden connections between the clustering of observation times and supernovae type, although it is hard to test for this.

The goal of the classifier is to determine the supernovae type in the test set. Two problems are considered, (1) to categorise two classes (type-Ia vs. non-type-Ia), and (2) to categorise three classes (supernovae types-1, -2 and -3). These are denoted 'SN1a' and '123' respectively. The first two problems are also attempted using only the first six observations with $S/N > 4$ and the data taken on the night of the sixth observation as described in [213].

Several metrics are used to assess the classifier. The simplest is the accuracy, defined as the ratio between the number of correct predictions and total number of predictions. With two classes a random classifier would have an accuracy of 0.5, and with three classes an accuracy of 1/3.

Next are a variety of metrics coming from the *confusion matrix* of predictions. For binary classification problems, the confusion matrix splits predictions into true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). We consider the purity and completeness of the classifier. These are defined as

$$\text{Purity} \quad = \quad \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{IV–2.2}$$

$$\text{Completeness} \quad = \quad \frac{\text{TP}}{\text{TP} + \text{FN}}\,. \tag{IV–2.3}$$

These are evaluated for each class separately vs. 'the rest' (e.g. type-Ia vs. non-type-Ia). The SPCC also defined the F_1 figure-of-merit for the SN1a classification problem. This is

$$F_1 \quad = \quad \frac{1}{\mathrm{TP}+\mathrm{FN}} \frac{\mathrm{TP}^2}{\mathrm{TP}+3\times\mathrm{FP}}\,, \tag{IV–2.4}$$

so incorrectly classifying a non-type-Ia supernovae as a type-Ia is penalised more heavily.

Finally, the area under the curve (AUC) is calculated. The AUC is the area under the curve of the TP rate vs. FP rate, as the threshold probability for classification is increased from 0 to 1. A perfect classifier has an AUC of 1, and a random classifier 0.5. For multi-class problems, the AUC is calculated for each class vs. the rest, and an unweighted average is taken to give the final AUC score.

2.2 Network architecture

Several combinations of the network architecture are taken. For the RNN type in the hidden layers, both vanilla RNN and long term memory (LSTM and GRU) units are tested. Both unidirectional and bidirectional networks are also attempted. For unidirectional networks the direction is fixed to be forwards. For bidirectional networks, the number of hidden units in each RNN layer is equal in the forward and backward directions.

Stacking two sets of layers is also tested to form a deep network. In the unidirectional case two hidden layers are stacked. In the bidirectional case the two stacks consists of a pair of forwards and backwards layers. The number of hidden units in a network with a single stack are denoted by $[h_1]$, and the number of hidden layers in a two stack model by $[h_1, h_2]$. The number of hidden units are varied, testing h = [4], [8], [16], [32], [4, 4], [8, 8], [16, 16] and [32, 32]. A stack of two layers is the maximum number tested due to the limited size of the data set.

For each network five randomised runs are performed over the training data to obtain the classifier metrics. The loss function is the categorical cross-entropy between the predictions and test data. The network weights are trained using back propagation with the `Adam` updater [221]. Mini-batches containing 10 samples are used throughout, and each model is trained for 200 *epochs*, where each epoch is a full pass over the training data. If training with a graphics processing unit (GPU), larger mini-batches are recommended to make use of the GPU

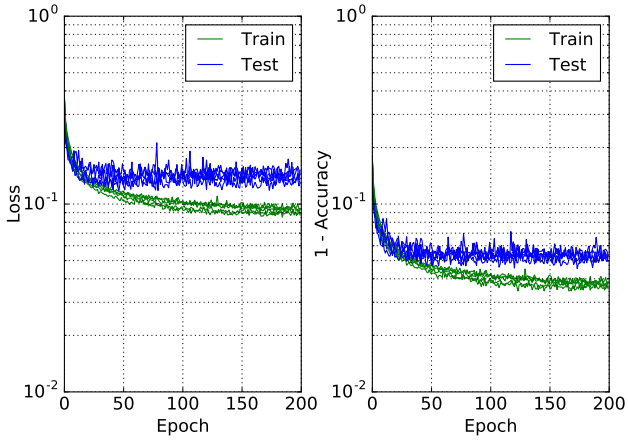


Figure IV–2.4: (Left) Training loss (green) vs. test loss (blue) for a unidirectional 2 layer LSTM network with 16 hidden units in each layer. (Right) Training accuracy (green) vs. test accuracy (blue) for the same network.

cores.

2.3 Results

A data set of 21,319 is relatively small by deep learning standards. Furthermore, the 'feature space' of supernovae light curves is significantly smaller than, say, using RNN to learn about language. Care therefore needs to be taken about *over-fitting*.

For a training fraction of 0.5, the best architecture was a deep 2-layer network with unidirectional LSTM units. Bidirectional units did not significantly improve the test accuracy and made the network more difficult to train. There was a marked improvement in test accuracy using 16 hidden units in each layer rather than 8, but too much over-fitting occurred using 32 hidden units. Over-fitting was still an issue for 16 hidden units, but a dropout could regularise this. As described in Part IV–1.7 dropout sets a random fraction of connections to 0 at each update during training *only*, preventing the units from adapting too much. Dropout is only applied to non-recurrent connections after each hidden layer.

Figure IV–2.4 shows the training and tests losses for such a network, with a dropout of 0.5, applied to type-Ia vs. non-type-Ia classification with host galaxy photo-z information. Without dropout the training loss continues to fall and the test loss rises. For five randomised runs, training for 200 epochs, a classification

accuracy of $94.9 \pm 0.2\%$, AUC of 0.986 ± 0.001 and $F_1 = 0.64 \pm 0.01$ are obtained. The corresponding type-Ia purity and completeness are $87.3 \pm 0.8\%$ and $91.4 \pm 1.1\%$ respectively. A summary of results and comparisons can be found in tables IV–2.2 and IV–2.3. The inclusion of host galaxy photo-z marginally improves the classifier performance. The 1σ errors quoted in the table are the result of five runs where the training data is randomly chosen (and so different) each time. Some random choice of the set of light curves are more effective for training the network than others, but it is extremely difficult to optimise this.

To test the robustness of the time-grouping method 10% of the known filter values (and/or their errors) are removed before grouping the data into a single vector and randomly augmenting the missing values. After training there is a small degradation in the results, i.e. for a training fraction of 0.5 using a deep two-layer, unidirectional network with 16 hidden units, a dropout of 0.5 and including the photo-z information the obtained results are very similar to the second line in table IV–2.2. This shows that a reduction in 10% of the points is similar to the omission of the photo-z data and therefore the data augmentation method is extremely robust.

One advantage of the approach used here is that light curve data can be directly input to a *pre-trained* model to give very fast evaluation (< 1s) of supernovae type. In the lower panel of figure IV–2.2 the light curve of a type-Ia supernovae (excluded from training) is inputted as a function of time to the pre-trained two-layer LSTM model discussed above. The classifier (type-Ia vs. non-type-Ia) is initially unsure of classification, with a type-Ia probability of around 0.5. The probability then decreases slightly, but rapidly increases near the peak of the light curve. The classifier has high confidence the supernovae is of type-Ia at around 60 days, and the final probability is excess of 99.5%. This method could therefore be useful to give early indication of supernovae type in surveys.

The same model is tested using a training fraction of 0.25 (around 5000 supernovae), closer to the lower end of the number likely to be followed up for LSST. After five randomised runs and training for 200 epochs an accuracy of $92.9 \pm 0.6\%$, AUC of 0.975 ± 0.003 and $F_1 = 0.57 \pm 0.03$ is obtained. The corresponding type-Ia purity and completeness are $86.6 \pm 2.0\%$ and $83.4 \pm 3.4\%$ respectively. The F_1 metric has degraded by $\sim 10\%$ for a reduction in data of 50%.

For 5.2% of the representative SPCC data, the training data set is so small

Method	Training size	AUC	Accuracy (%)	F_1	Purity (%)	Completeness (%)	Photo-z
A	10,660	0.986 ± 0.001	94.7 ± 0.2	0.64 ± 0.01	87.3 ± 0.8	91.4 ± 1.1	True
A	10,660	0.981 ± 0.001	93.6 ± 0.3	0.60 ± 0.02	87.4 ± 1.7	85.4 ± 2.6	False
A	5,330	0.975 ± 0.003	92.9 ± 0.6	0.57 ± 0.03	86.6 ± 2.0	83.4 ± 3.4	True
A	5,330	0.973 ± 0.002	92.3 ± 0.4	0.55 ± 0.02	86.2 ± 2.4	80.8 ± 3.8	False
B	1,103	0.910 ± 0.012	85.9 ± 0.9	0.31 ± 0.03	72.4 ± 0.4	66.1 ± 6.0	True
B	1,103	0.901 ± 0.016	84.6 ± 1.7	0.28 ± 0.05	68.2 ± 3.4	66.3 ± 5.5	False
C	∼10,660	-	-	0.58	85	88	True
C	∼10,660	-	-	0.51	82	85	False
C	1,045	-	-	0.33	70	75	True
C	1,045	-	-	0.29	67	71	False
D	∼8,000	-	-	0.55	-	-	True
D	∼2,000	-	-	0.45	-	-	True
E	1,103	0.94 ± 0.03	-	-	-	-	True
E	1,103	0.89 ± 0.53	-	-	-	-	False
E	1,103	-	-	-	90	85	True
E	1,103	-	-	-	87	90	True

Table IV–2.2: Summary of results for type-Ia vs. non-type-Ia classification with a training fraction of 0.5, 0.25 and 0.052 with comparisons to similar methods in [208] and [269]. The models used are A) Unidirectional LSTM, [16, 16] with 0.5 dropout, B) Unidirectional LSTM, [4] with 0.5 dropout, C) [208], D) [269], E) [242] SALT2 fits averaged over machine learning architecture. Errors on results are the mean and standard deviation values from five randomised runs.

that over-fitting is more severe. Using the same two-layer LSTM network with sixteen hidden units and dropout of 0.5 there is a notable increase in the test loss after $\sim$ 20 epochs, but the accuracy and other metrics remain relatively constant (F_1 values of 0.35 to 0.4 were obtained). The reason for this apparent discrepancy is that the accuracy, say, simply takes the maximum value of the `softmax` output layer. For example, a two-class problem with output probabilities [0.6, 0.4] and target [1, 0] has the same accuracy as one with output probabilities [0.8, 0.2]. The loss in the latter case would be lower however, and represents increased confidence of the network in its predictions. Models with severe over-fitting and an increasing cross-entropy loss at the expense of metrics such as F_1 are therefore rejected and the model complexity is decreased.

For a training fraction of 5.2% a single-layer LSTM network, with 4 hidden units, and dropout of 0.5 satisfies this criteria. For five randomised runs, training for 200 epochs, a classification accuracy of $85.9 \pm 0.9\%$, AUC of 0.910 ± 0.012 and $F_1 = 0.31 \pm 0.03$ is obtained. The corresponding type-Ia purity and completeness are $72.4 \pm 0.4\%$ and $66.1 \pm 6.0\%$ respectively.

It is difficult to directly compare the results from the SPCC challenge in [212] with this work since the figure of merit is quoted as a function of redshift and a non-representative set of light curves was originally used. In [212] the method of [314] had the highest average F_1, with 79% purity and 96% accuracy. This is a, somewhat, confusing average as $F_1 \sim 0.4$ at a redshift $z \sim 0.1$ up to $F_1 \sim 1$ at $z \sim 0.9$. Other methods performed similarly.

It is better to consider comparison with other methods using post-SPCC data. The results obtained here are competitive with previous approaches. The analyses by [208] and [269] are easier to compare. Along with [242] these employ a two-step process, where features are first extracted by various methods before machine learning classification. The results obtained for similar sized training sets are comparable as can be seen in the top section of table IV–2.2. When using half the data set to train on we get a higher F_1 value, $F_1 = 0.64$ compared to $F_1 = 0.58$ in [208]. The value in [269] is also similar given that the sample size is smaller. For a smaller sample training set of 5.2% of all the data we again perform similarly to [208] but under perform compare to [269] taking into account the slightly larger sample size in the latter case. In [242] using the SALT2 fits provided the best average AUC over a range of machine learning techniques. By imposing a purity of 90% a completeness of 85% was achieved while requiring a completeness of 90% reveals a corresponding purity of 85%.

The top section of table IV–2.3 shows the results of the early-epoch challenge from SPCC. Here only the data before the night of the sixth observation with $S/N > 4$ for each light curve can be used - a great reduction from the use of the full light curve. An accuracy of $93.1 \pm 0.4\%$, AUC of 0.977 ± 0.002 and an $F_1 = 0.58 \pm 0.01$ with a training fraction of 0.5 and including host-z gives surprisingly good results. These values are not far from those obtained using the whole light curve and are equivalent to the full results of [208]. The results are not as good with a training fraction of 0.052, but still comparable to the results found here using the whole light curve. The network trained on the partial light curves does better than suggested from feeding the early-epoch light curve through a network trained on the full sequence. This is due to the later parts of the light curve influencing the weights of the network whilst training. Training on only the initial part of the light curve optimises the network weights such that early sequence features have more effect, resulting in better accuracy, AUC and F_1 values than expected.

In the middle section of table IV–2.3 the three class categorisation is shown. There is no available data for comparison of this problem, but compared to classification between type-Ia vs. non-type-Ia, bidirectional RNN do well. The AUC and accuracy remain high, still above 90% when the host-z is included using a training fraction of 0.5. Using a smaller training fraction of 0.052, the results are worsened similar to the two class categorisation in table IV–2.2.

Finally, the bottom section of table IV–2.3 has the results of the three class categorisation when using the early-epoch data. The results are similar to the difference between the full light curve and early-epoch data SN1a categorisation when comparing with the full light curve 123 categorisation. It should be noted that the bidirectional network used for the 123 categorisation using the full light curve revealed sizeable over-fitting when using the early-epoch data and so a unidirectional network was used instead.

Method	Training size	AUC	Accuracy (%)	F_1	Purity (%)	Completeness (%)	Photo-z
A	10,660	0.977 ± 0.002	93.1 ± 0.4	0.58 ± 0.01	88.0 ± 1.1	82.2 ± 2.8	True
A	10,660	0.970 ± 0.001	92.0 ± 0.3	0.53 ± 0.01	86.0 ± 0.9	79.5 ± 2.2	False
B	1,103	0.902 ± 0.014	85.2 ± 1.2	0.29 ± 0.04	71.5 ± 1.6	62.8 ± 5.6	True
B	1,103	0.860 ± 0.017	81.6 ± 1.2	0.21 ± 0.02	62.6 ± 3.0	57.6 ± 2.7	False
F	10,660	0.974 ± 0.001	90.4 ± 0.3	-	90.6 ± 0.7	86.5 ± 0.7	True
F	10,660	0.959 ± 0.006	88.5 ± 1.1	-	87.6 ± 1.1	85.9 ± 4.1	False
G	1,103	0.868 ± 0.015	78.1 ± 0.9	-	70.8 ± 3.4	70.6 ± 4.1	True
G	1,103	0.865 ± 0.011	78.0 ± 1.2	-	66.9 ± 3.2	74.5 ± 4.2	False
A	10,660	0.960 ± 0.006	87.9 ± 0.9	-	86.4 ± 0.8	84.4 ± 3.5	True
A	10,660	0.948 ± 0.002	86.8 ± 0.3	-	84.1 ± 1.1	83.7 ± 1.4	False
B	1,103	0.851 ± 0.013	76.8 ± 1.3	-	64.7 ± 3.8	71.0 ± 4.1	True
B	1,103	0.819 ± 0.010	74.2 ± 1.0	-	58.1 ± 3.8	73.6 ± 6.6	False

Table IV–2.3: (Top section) Summary of results for SPCC early-epoch challenge. (Middle section) Summary of results for types-I, -II and -III classification. (Bottom section) Summary of the results for the SPCC early-epoch challenge when classifying between types-I , -II and -III supernovae. The models used are A) Unidirectional LSTM, [16, 16] with 0.5 dropout, B) Unidirectional LSTM, [4] with 0.5 dropout (both as in table IV–2.2), F) Bidirectional LSTM, [16, 16] with 0.5 dropout, G) Bidirectional LSTM, [4] with 0.5 dropout. Errors on results are the mean and standard deviation values from five randomised runs.

Chapter 3

Discussion

A new method for performing photometric classification of supernovae has been presented in this Part. Machine learning methodology has previously been applied to SPCC classification [208; 242; 269]. Instead of performing feature extraction before classification, the approach here uses the light curves directly as inputs to an RNN, which is able to learn information from the sequence of observations.

Although the network has been trained on the cross-entropy loss and not the F_1 score, for the same sized data set of $\sim 10^3(10^4)$ supernovae (including host galaxy photo-z), the method in [208] obtained F_1 values of 0.33 (0.58), and the method in [269] got values of 0.42 (0.57), compared to 0.31 (0.64) here. RNN therefore compare well with other methods when a larger training set is available. The performance isn't quite as good with a smaller training set, possibly due to the network having to learn from no prior information about (noisy) light curves. The current state-of-the-art for a small training set ($\sim 10^3$ supernovae) comes from a combination of SALT2 (Spectral Adaptive Light curve Template 2) template fits and boosted decision trees [242]. It would be interesting to check how deep learning compares to this with a larger training set.

As well as finding competitive results for the final metrics, it is possible to give fast, early evaluation of supernovae type using pre-trained models. This is possible since the light curve can be fed to the model directly without needing any feature extraction. Most interestingly, training a network only on the early epoch light curve data results in a better early-time predictor than using a network trained on entire light curve data. The results using only the early-epoch data are close to those using the entire light curve data for both SN1a and 123 categorisation with both large and small training fractions.

There are several possibilities for future work. One of the advantages of RNN are that inputs are agnostic, so the impact of any additional inputs could be explored. It would be possible, for example, to even pass the raw images in each filter though a convolutional network and use those as inputs. A representative training sample has been considered here, but spectroscopic follow up surveys may be biased. The performance of the network could be measured against selection bias, and the results used to inform the best follow up strategy. Further work could also be performed to optimise the early detection probability of the network. Finally, to improve performance in the small data regime one can use *transfer learning.* Here, a more complex network is pre-trained on simulations or existing data from other surveys, then the weights of the network are fine-tuned on the new, smaller data set.

Other than using RNN to learn about time domain physics, a very exciting era of machine learning in cosmology and astronomy is on the horizon. The use of convolutional neural networks to data mine surveys will probe hidden features which are missed using conventional techniques [211]. Belief networks [303] will also allow statistical interpretation of results to be processed without human bias, relating to the outcomes of future work from Part III. As greater amounts of data are obtained from a variety of different experiments then work on learning how to amalgamate different networks to cope with different input types is necessary [146]. Outside of the cosmological community, machine learning can almost certainly be argued to be the most popular and thriving collective in computing today. Every problem which has a large data set is most likely being tackled with deep learning, even now. Where there is not a large data set, neural networks are being trained using unsupervised learning or reinforcement learning. The field of machine learning is burgeoning and many tasks in the near future will undoubtedly prosper due to its growth.

Bibliography

[1] T. Abbott et al. "Cosmology from cosmic shear with Dark Energy Survey Science Verification data". In: *Physical Review* D94.2 (2016), p. 022001. DOI: 10.1103/PhysRevD.94.022001. arXiv: 1507.05552 [astro-ph.CO].

[2] P. A. Abell et al. "LSST Science Book, Version 2.0". In: *ArXiv e-prints* (Dec. 2009). arXiv: 0912.0201 [astro-ph.IM].

[3] A. A. Abrikosov. "On the Magnetic properties of superconductors of the second group". In: *Soviet Physics – Journal of Experimental and Theoretical Physics* 5 (1957). [Zh. Eksp. Teor. Fiz.32,1442(1957)], pp. 1174–1182.

[4] R. Adam et al. "Planck intermediate results. XLVII. Planck constraints on reionization history". In: *Astronomy & Astrophysics* 596 (2016), A108. DOI: 10.1051/0004-6361/201628897. arXiv: 1605.03507 [astro-ph.CO].

[5] G. E. Addison et al. "Quantifying discordance in the 2015 Planck CMB spectrum". In: *The Astrophysical Journal* 818.2 (2016), p. 132. DOI: 10.3847/0004-637X/818/2/132. arXiv: 1511.00055 [astro-ph.CO].

[6] P. A. R. Ade et al. "Joint Analysis of BICEP2/*Keck Array* and *Planck* Data". In: *Physical Review Letters* 114 (2015), p. 101301. DOI: 10.1103/PhysRevLett.114.101301. arXiv: 1502.00612 [astro-ph.CO].

[7] P. A. R. Ade et al. "Planck 2013 results. I. Overview of products and scientific results". In: *Astronomy & Astrophysics* 571 (2014), A1. DOI: 10.1051/0004-6361/201321529. arXiv: 1303.5062 [astro-ph.CO].

[8] P. A. R. Ade et al. "Planck 2013 results. XV. CMB power spectra and likelihood". In: *Astronomy & Astrophysics* 571 (2014), A15. DOI: 10.1051/0004-6361/201321573. arXiv: 1303.5075 [astro-ph.CO].

[9] P. A. R. Ade et al. "Planck 2013 results. XVI. Cosmological parameters". In: *Astronomy & Astrophysics* 571 (2014), A16. DOI: 10.1051/0004-6361/201321591. arXiv: 1303.5076 [astro-ph.CO].

[10] P. A. R. Ade et al. "Planck 2013 results. XVII. Gravitational lensing by large-scale structure". In: *Astronomy & Astrophysics* 571 (2014), A17. DOI: 10.1051/0004-6361/201321543. arXiv: 1303.5077 [astro-ph.CO].

[11] P. A. R. Ade et al. "Planck 2013 results. XVIII. The gravitational lensing-infrared background correlation". In: *Astronomy & Astrophysics* 571 (2014), A18. DOI: 10.1051/0004-6361/201321540. arXiv: 1303.5078.

[12] P. A. R. Ade et al. "Planck 2013 results. XX. Cosmology from Sunyaev–Zeldovich cluster counts". In: *Astronomy & Astrophysics* 571 (2014), A20. DOI: 10.1051/0004-6361/201321521. arXiv: 1303.5080 [astro-ph.CO].

[13] P. A. R. Ade et al. "Planck 2013 results. XXV. Searches for cosmic strings and other topological defects". In: *Astronomy & Astrophysics* 571 (2014), A25. DOI: 10.1051/0004-6361/201321621. arXiv: 1303.5085 [astro-ph.CO].

[14] P. A. R. Ade et al. "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy & Astrophysics* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].

[15] P. A. R. Ade et al. "Planck 2015 results. XV. Gravitational lensing". In: *Astronomy & Astrophysics* 594 (2016), A15. DOI: 10.1051/0004-6361/201525941. arXiv: 1502.01591 [astro-ph.CO].

[16] P. A. R. Ade et al. "Planck 2015 results. XX. Constraints on inflation". In: *Astronomy & Astrophysics* 594 (2016), A20. DOI: 10.1051/0004-6361/201525898. arXiv: 1502.02114 [astro-ph.CO].

[17] P. A. R. Ade et al. "Planck 2015 results. XXIV. Cosmology from Sunyaev–Zeldovich cluster counts". In: *Astronomy & Astrophysics* 594 (2016), A24. DOI: 10.1051/0004-6361/201525833. arXiv: 1502.01597 [astro-ph.CO].

[18] N. Aghanim et al. "Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters". In: *Astronomy & Astrophysics* 594 (2016), A11. DOI: 10.1051/0004-6361/201526926. arXiv: 1507.02704 [astro-ph.CO].

[19] N. Aghanim et al. "Planck 2016 intermediate results. LI. Features in the cosmic microwave background temperature power spectrum and shifts in cosmological parameters". In: (2016). arXiv: 1608.02487 [astro-ph.CO].

[20] A. Aguilar et al. "Evidence for neutrino oscillations from the observation of anti-neutrino(electron) appearance in a anti-neutrino(muon) beam". In: *Physical Review* D64 (2001), p. 112007. DOI: 10.1103/PhysRevD.64.112007. arXiv: hep-ex/0104049 [hep-ex].

[21] A. A. Aguilar-Arevalo et al. "The Neutrino Flux prediction at MiniBooNE". In: *Physical Review* D79 (2009), p. 072002. DOI: 10.1103/PhysRevD.79.072002. arXiv: 0806.1449 [hep-ex].

[22] A. Albrecht, R. A. Battye, and J. Robinson. "Detailed study of defect models for cosmic structure formation". In: *Physical Review* D59 (1998), p. 023508. DOI: 10.1103/PhysRevD.59.023508. arXiv: astro-ph/9711121 [astro-ph].

[23] A. Albrecht, R. A. Battye, and J. Robinson. "The Case against scaling defect models of cosmic structure formation". In: *Physical Review Letters* 79 (1997), pp. 4736–4739. DOI: 10.1103/PhysRevLett.79.4736. arXiv: astro-ph/9707129 [astro-ph].

[24] A. Albrecht and N. Turok. "Evolution of cosmic string networks". In: *Physical Review* D40 (4 1989), pp. 973–1001. DOI: 10.1103/PhysRevD.40.973.

[25] B. Allen and E. P. S. Shellard. "Cosmic string evolution: a numerical simulation". In: *Physical Review Letters* 64 (1990), pp. 119–122. DOI: 10.1103/PhysRevLett.64.119.

[26] R. A. Alpher and R. C. Herman. "On the Relative Abundance of the Elements". In: *Physical Review* 74 (12 1948), pp. 1737–1742. DOI: 10.1103/PhysRev.74.1737.

[27] L. Anderson et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample". In: *Monthly Notices of the Royal Astronomical Society* 427.4 (2013), pp. 3435–3467. DOI: 10.1111/j.1365-2966.2012.22066.x. arXiv: 1203.6594 [astro-ph.CO].

[28] M. Archidiacono et al. "Sterile neutrinos: Cosmology versus short-baseline experiments". In: *Physical Review* D87.12 (2013), p. 125034. DOI: 10.1103/PhysRevD.87.125034. arXiv: 1302.6720 [astro-ph.CO].

[29] B. Armbruster et al. "Upper limits for neutrino oscillations muon-anti-neutrino —> electron-anti-neutrino from muon decay at rest". In: *Physical Review* D65 (2002), p. 112001. DOI: 10.1103/PhysRevD.65.112001. arXiv: hep-ex/0203021 [hep-ex].

[30] L. Auslander and R. E. MacKenzie. *Introduction to Differentiable Manifolds*. Dover Books on Mathematics. Dover Publications, 2012.

[31] D. Austin, E. J. Copeland, and R. J. Rivers. "Statistical mechanics of strings on periodic lattices". In: *Physical Review* D49 (8 1994), pp. 4089–4093. DOI: 10.1103/PhysRevD.49.4089.

[32] P. P. Avelino et al. "Cosmic string seeded structure formation". In: *Physical Review Letters* 81 (1998), pp. 2008–2011. DOI: 10.1103/PhysRevLett.81.2008. arXiv: astro-ph/9712008 [astro-ph].

[33] A. Avgoustidis and E. J. Copeland. "The effect of kinematic constraints on multi-tension string network evolution". In: *Physical Review* D81 (2010), p. 063517. DOI: 10.1103/PhysRevD.81.063517. arXiv: 0912.4004 [hep-ph].

[34] A. Avgoustidis and E. P. S. Shellard. "Effect of reconnection probability on cosmic (super)string network density". In: *Physical Review* D73 (2006), p. 041301. DOI: 10.1103/PhysRevD.73.041301. arXiv: astro-ph/0512582 [astro-ph].

[35] A. Avgoustidis and E. P. S. Shellard. "Velocity-Dependent Models for Non-Abelian/Entangled String Networks". In: *Physical Review* D78 (2008), p. 103510. DOI: 10.1103/PhysRevD.78.10351010.1103/PhysRevD.80.129907,10.1103/PhysRevD.78.103510,10.1103/PhysRevD.80.129907. arXiv: 0705.3395 [astro-ph].

[36] A. Avgoustidis et al. "Constraints on the fundamental string coupling from B-mode experiments". In: *Physical Review Letters* 107 (2011), p. 121301. DOI: 10.1103/PhysRevLett.107.121301. arXiv: 1105.6198 [astro-ph.CO].

[37] A. Avgoustidis et al. "Fast Analytic Computation of Cosmic String Power Spectra". In: *Physical Review* D86 (2012), p. 123513. DOI: 10.1103/PhysRevD.86.123513. arXiv: 1209.2461 [astro-ph.CO].

[38] T. Baldauf et al. "Evidence for quadratic tidal tensor bias from the halo bispectrum". In: *Physical Review* 8, 083540 (Oct. 2012), p. 083540. DOI: 10.1103/PhysRevD.86.083540. arXiv: 1201.4827 [astro-ph.CO].

[39] N. M. Ball and R. J. Brunner. "Data Mining and Machine Learning in Astronomy". In: *International Journal of Modern Physics* (2010), pp. 1049–1106. DOI: 10.1142/S0218271810017160. arXiv: 0906.2173 [astro-ph.IM].

[40] N. M. Ball et al. "Robust Machine Learning Applied to Astronomical Datasets. 1. Star-Galaxy Classification of the SDSS DR3 Using Decision Trees". In: *The Astrophysical Journal* 650 (2006), pp. 497–509. DOI: 10.1086/507440. arXiv: astro-ph/0606541 [astro-ph].

[41] W. E. Ballinger, J. A. Peacock, and A. F. Heavens. "Measuring the cosmological constant with redshift surveys". In: *Monthly Notices of the Royal Astronomical Society* 282 (1996), pp. 877–888. DOI: 10.1093/mnras/282.3.877. arXiv: astro-ph/9605017 [astro-ph].

[42] M. Banerji et al. "Galaxy Zoo: Reproducing Galaxy Morphologies Via Machine Learning". In: *Monthly Notices of the Royal Astronomical Society* 406 (2010), p. 342. DOI: 10.1111/j.1365-2966.2010.16713.x. arXiv: 0908.2033 [astro-ph.CO].

[43] J. Barriga et al. "On the APM power spectrum and the CMB anisotropy: Evidence for a phase transition during inflation?" In: *Monthly Notices of the Royal Astronomical Society* 324 (2001), p. 977. DOI: 10.1046/j.1365-8711.2001.04373.x. arXiv: astro-ph/0011398 [astro-ph].

[44] R. A. Battye, T. Charnock, and A. Moss. "Tension between the power spectrum of density perturbations measured on large and small scales". In: *Physical Review* D91.10 (2015), p. 103508. DOI: 10.1103/PhysRevD.91.103508. arXiv: 1409.2769 [astro-ph.CO].

[45] R. A. Battye and A. Moss. "Evidence for massive neutrinos from CMB and lensing observations". In: *Physical Review Letters* 112 (2014), p. 051303. DOI: 10.1103/PhysRevLett.112.051303. arXiv: 1308.5870.

[46] R. A. Battye and A. Moss. "Updated constraints on the cosmic string tension". In: *Physical Review* D82 (2010), p. 023521. DOI: 10.1103/PhysRevD.82.023521. arXiv: 1005.0479 [astro-ph.CO].

[47] R. A. Battye, J. Robinson, and A. Albrecht. "Structure formation by cosmic strings with a cosmological constant". In: *Physical Review Letters* 80 (1998), pp. 4847–4850. DOI: 10.1103/PhysRevLett.80.4847. arXiv: astro-ph/9711336 [astro-ph].

[48] D. Baumann et al. "Towards an Explicit Model of D-brane Inflation". In: *Journal of Cosmology and Astroparticle Physics* 0801 (2008), p. 024. DOI: 10.1088/1475-7516/2008/01/024. arXiv: 0706.0360 [hep-th].

[49] R. H. Becker et al. "Evidence for reionization at Z 6: Detection of a Gunn-Peterson trough in a Z = 6.28 quasar". In: *The Astronomical Journal* 122 (2001), p. 2850. DOI: 10.1086/324231. arXiv: astro-ph/0108097 [astro-ph].

[50] M. Benetti. "Updating constraints on inflationary features in the primordial power spectrum with the Planck data". In: *Physical Review* D88 (2013), p. 087302. DOI: 10.1103/PhysRevD.88.087302. arXiv: 1308.6406 [astro-ph.CO].

[51] M. Benetti et al. "Featuring the primordial power spectrum: new constraints on interrupted slow-roll from CMB and LRG data". In: *Physical Review* D87 (2013), p. 023519. DOI: 10.1103/PhysRevD.87.023519. arXiv: 1210.3562 [astro-ph.CO].

[52] Y. Bengio. "Learning Deep Architectures for AI". In: *Foundations and Trends in Machine Learning* 2.1 (2009), pp. 1–127. ISSN: 1935-8237. DOI: 10.1561/2200000006.

[53] C. L. Bennett et al. "The Microwave Anisotropy Probe Mission". In: *The Astrophysical Journal* 583 (Jan. 2003), pp. 1–23. DOI: 10.1086/345346. eprint: astro-ph/0301158.

[54] D. P. Bennett and F. R. Bouchet. "High resolution simulations of cosmic string evolution. 1. Network evolution". In: *Physical Review* D41 (1990), p. 2408. DOI: 10.1103/PhysRevD.41.2408.

[55] C. W. Bernard. "Feynman rules for gauge theories at finite temperature". In: *Physical Review* D9 (12 1974), pp. 3312–3320. DOI: 10.1103/PhysRevD.9.3312.

[56] F. Beutler et al. "The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant". In: *Monthly Notices of the Royal Astronomical Society* 416 (2011), pp. 3017–3032. DOI: 10.1111/j.1365-2966.2011.19250.x. arXiv: 1106.3366 [astro-ph.CO].

[57] F. Beutler et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Signs of neutrino mass in current cosmological datasets". In: *Monthly Notices of the Royal Astronomical Society* 444.4 (2014), pp. 3501–3516–3516. DOI: 10.1093/mnras/stu1702. arXiv: 1403.4599 [astro-ph.CO].

[58] F. Beutler et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles". In: *Monthly Notices of the Royal Astronomical Society* 443.2 (2014), pp. 1065–1089. DOI: 10.1093/mnras/stu1051. arXiv: 1312.4611 [astro-ph.CO].

[59] N. Bevis et al. "CMB polarization power spectra contributions from a network of cosmic strings". In: *Physical Review* D76 (2007), p. 043005. DOI: 10.1103/PhysRevD.76.043005. arXiv: 0704.3800 [astro-ph].

[60] N. Bevis et al. "CMB power spectra from cosmic strings: predictions for the Planck satellite and beyond". In: *Physical Review* D82 (2010), p. 065004. DOI: 10.1103/PhysRevD.82.065004. arXiv: 1005.2663 [astro-ph.CO].

[61] A. Bhattacharyya. "On a measure of divergence between two statistical populations defined by their probability distributions". In: *Bulletin of Calcutta Mathematical Society* 35 (1943), pp. 99–109.

[62] S. Bird, M. Viel, and M. G. Haehnelt. "Massive Neutrinos and the Non-linear Matter Power Spectrum". In: *Monthly Notices of the Royal Astronomical Society* 420 (2012), pp. 2551–2561. DOI: 10.1111/j.1365-2966.2011.20222.x. arXiv: 1109.4416 [astro-ph.CO].

[63] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer. "Large parallel cosmic string simulations: New results on loop production". In: *Physical Review* D83 (2011), p. 083514. DOI: 10.1103/PhysRevD.83.083514. arXiv: 1101.5173 [astro-ph.CO].

[64] J. S. Bloom and J. W. Richards. "Data Mining and Machine Learning in Time-Domain Discovery and Classification". In: *Advances in Machine Learning and Data Mining for Astronomy*. Ed. by M. J. Way et al. Mar. 2012, pp. 89–112.

[65] G. R. Blumenthal, H. Pagels, and J. R. Primack. "Galaxy formation by dissipationless particles heavier than neutrinos". In: *Nature* 299 (Sept. 1982), p. 37. DOI: 10.1038/299037a0.

[66] G. R. Blumenthal et al. "Formation of galaxies and large-scale structure with cold dark matter". In: *Nature* 311 (Oct. 1984), pp. 517–525. DOI: 10.1038/311517a0.

[67] J. R. Bond and G. Efstathiou. "Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter". In: *The Astrophysical Journal Letters* 285 (Oct. 1984), pp. L45–L48. DOI: 10.1086/184362.

[68] J. R. Bond and G. Efstathiou. "The statistics of cosmic background radiation fluctuations". In: *Monthly Notices of the Royal Astronomical Society* 226 (June 1987), pp. 655–687. DOI: 10.1093/mnras/226.3.655.

[69] J. R. Bond, A. S. Szalay, and M. S. Turner. "Formation of galaxies in a gravitino-dominated universe". In: *Physical Review Letters* 48 (June 1982), pp. 1636–1639. DOI: 10.1103/PhysRevLett.48.1636.

[70] H. Bourlard and Y. Kamp. "Auto-association by multilayer perceptrons and singular value decomposition". In: *Biological Cybernetics* 59.4 (1988), pp. 291–294. ISSN: 1432-0770. DOI: 10.1007/BF00332918.

[71] R. H. Brandenberger. "Cosmic strings and the large scale structure of the universe". In: *Physica Scripta* T36 (1991), pp. 114–126. DOI: 10.1088/0031-8949/1991/T36/013.

[72] R. H. Brandenberger. *Large Scale Structure Formation*. Springer Science & Business Media, 2000, pp. 296–298.

[73] S. L. Bridle et al. "Cosmological parameters from cluster abundances, cosmic microwave background and IRAS". In: *Monthly Notices of the Royal Astronomical Society* 310 (Dec. 1999), pp. 565–570. DOI: 10.1046/j.1365-8711.1999.02960.x. eprint: astro-ph/9903472.

[74] C. J. Burden. "Gravitational radiation from a particular class of cosmic strings". In: *Physics Letters* B164.4 (1985), pp. 277 –281. ISSN: 0370-2693. DOI: 10.1016/0370-2693(85)90326-0.

[75] C. P. Burgess et al. "The Inflationary brane anti-brane universe". In: *Journal of High Energy Physics* 0107 (2001), p. 047. DOI: 10.1088/1126-6708/2001/07/047. arXiv: hep-th/0105204 [hep-th].

[76] E. Calabrese et al. "Cosmic Microwave Weak lensing data as a test for the dark universe". In: *Physical Review* D77 (2008), p. 123531. DOI: 10.1103/PhysRevD.77.123531. arXiv: 0803.2309 [astro-ph].

[77] L. Campanelli et al. "Galaxy cluster number count data constraints on cosmological parameters". In: *European Physical Journal*, 2218 (Nov. 2012), p. 2218. DOI: 10.1140/epjc/s10052-012-2218-4. arXiv: 1110.2310.

[78] F. Capozzi et al. "Status of three-neutrino oscillation parameters, circa 2013". In: *Physical Review* D89 (2014), p. 093018. DOI: 10.1103/PhysRevD.89.093018. arXiv: 1312.2878 [hep-ph].

[79] B. Carter. "Duality relation between charged elastic strings and superconducting cosmic strings". In: *Physics Letters* B224.1 (1989), pp. 61 –66. ISSN: 0370-2693. DOI: 10.1016/0370-2693(89)91051-4.

[80] K. C. Chan, R. Scoccimarro, and R. K. Sheth. "Gravity and Large-Scale Non-local Bias". In: *Physical Review* D85 (2012), p. 083509. DOI: 10.1103/PhysRevD.85.083509. arXiv: 1201.3614 [astro-ph.CO].

[81] D. Chandler. "The norm of the Schur product operation." In: *Numerische Mathematik* 4.1 (1962), pp. 343–344.

[82] T. Charnock et al. "Something to do with the cosmic string bispectrum and/or trispectrum". In: *In prep.* (2017).

[83] F. Chollet. *Keras RNN*. https://keras.io/layers/recurrent/. Keras documentation. 2014.

[84] J. Chung et al. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling". In: *ArXiv e-prints* (Dec. 2014). arXiv: 1412.3555.

[85] C. Contaldi, M. Hindmarsh, and J. Magueijo. "The Power spectra of CMB and density fluctuations seeded by local cosmic strings". In: *Physical Review Letters* 82 (1999), pp. 679–682. DOI: 10.1103/PhysRevLett.82.679. arXiv: astro-ph/9808201 [astro-ph].

[86] E. J. Copeland and T. W. B. Kibble. "Cosmic Strings and Superstrings". In: *Proceedings of the Royal Society of London* A466 (2010), pp. 623–657. DOI: 10.1098/rspa.2009.0591. arXiv: 0911.1345 [hep-th].

[87] E. J. Copeland, T. W. B. Kibble, and D. Austin. "Scaling solutions in cosmic-string networks". In: *Physical Review* D45 (4 1992), R1000–R1004. DOI: 10.1103/PhysRevD.45.R1000.

[88] E. J. Copeland, T. W. B. Kibble, and D. A. Steer. "Constraints on string networks with junctions". In: *Physical Review* D75 (2007), p. 065024. DOI: 10.1103/PhysRevD.75.065024. arXiv: hep-th/0611243 [hep-th].

[89] E. J. Copeland, J. Magueijo, and D. A. Steer. "Cosmological parameter dependence in local string theories of structure formation". In: *Physical Review* D61 (2000), p. 063505. DOI: 10.1103/PhysRevD.61.063505. arXiv: astro-ph/9903174 [astro-ph].

[90] E. J. Copeland, R. C. Myers, and J. Polchinski. "Cosmic F and D strings". In: *Journal of High Energy Physics* 0406 (2004), p. 013. DOI: 10.1088/1126-6708/2004/06/013. arXiv: hep-th/0312067 [hep-th].

[91] E. J. Copeland, L. Pogosian, and T. Vachaspati. "Seeking String Theory in the Cosmos". In: *Classical and Quantum Gravity* 28 (2011), p. 204009. DOI: 10.1088/0264-9381/28/20/204009. arXiv: 1105.0207 [hep-th].

[92] E. J. Copeland and P. M. Saffin. "On the evolution of cosmic-superstring networks". In: *Journal of High Energy Physics* 11 (2005), p. 023. DOI: 10.1088/1126-6708/2005/11/023. arXiv: hep-th/0505110 [hep-th].

[93] E. J. Copeland et al. "False vacuum inflation with Einstein gravity". In: *Physical Review* D49 (1994), pp. 6410–6433. DOI: 10.1103/PhysRevD.49.6410. arXiv: astro-ph/9401011 [astro-ph].

[94] E. J. Copeland et al. "On the Collision of Cosmic Superstrings". In: *Physical Review* D77 (2008), p. 063521. DOI: 10.1103/PhysRevD.77.063521. arXiv: 0712.0808 [hep-th].

[95] N. J. Cornish. "Using the acoustic peak to measure cosmological parameters". In: *Physical Review* D63 (2001), p. 027302. DOI: 10.1103/PhysRevD.63.027302. arXiv: astro-ph/0005261 [astro-ph].

[96] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Wiley, 2008.

[97] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals and Systems* 2.4 (1989), pp. 303–314. ISSN: 1435-568X. DOI: 10.1007/BF02551274.

[98] T. Damour and A. Vilenkin. "Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows". In: *Physical Review* D71 (2005), p. 063510. DOI: 10.1103/PhysRevD.71.063510. arXiv: hep-th/0410222 [hep-th].

[99] R. J. Danos, R. H. Brandenberger, and G. Holder. "A Signature of Cosmic Strings Wakes in the CMB Polarization". In: *Physical Review* D82 (2010), p. 023513. DOI: 10.1103/PhysRevD.82.023513. arXiv: 1003.0905 [astro-ph.CO].

[100] S. Das et al. "The Atacama Cosmology Telescope: temperature and gravitational lensing power spectrum measurements from three seasons of data". In: *Journal of Cosmology and Astroparticle Physics* 1404 (2014), p. 014. DOI: 10.1088/1475-7516/2014/04/014. arXiv: 1301.1037 [astro-ph.CO].

[101] D. Daverio et al. "Energy-momentum correlations for Abelian Higgs cosmic strings". In: *Physical Review* D93.8 (2016), p. 085014. DOI: 10.1103/PhysRevD.93.085014. arXiv: 1510.05006 [astro-ph.CO].

[102] Y. Declais et al. "Search for neutrino oscillations at 15-meters, 40-meters, and 95-meters from a nuclear power reactor at Bugey". In: *Nuclear Physics* B434 (1995), pp. 503–534. DOI: 10.1016/0550-3213(94)00513-E.

[103] A. Del Campo and W. H. Zurek. "Universality of phase transition dynamics: Topological defects from symmetry breaking". In: *International Journal of Modern Physics*, 1430018 (Mar. 2014), p. 1430018. DOI: 10.1142/S0217751X1430018X. arXiv: 1310.1600 [cond-mat.stat-mech].

[104] L. Deng and D. Yu. "Deep Learning: Methods and Applications". In: *Foundations and Trends in Signal Processing* 7.3–4 (2014), pp. 197–387. ISSN: 1932-8346. DOI: 10.1561/2000000039.

[105] S. Dimopoulos, J. Preskill, and F. Wilczek. "Catalyzed nucleon decay in neutron stars". In: *Physics Letters* B119.4 (1982), pp. 320 –322. ISSN: 0370-2693. DOI: 10.1016/0370-2693(82)90679-7.

[106] P. A. M. Dirac. "Quantised Singularities in the Electromagnetic Field". In: *Proceedings of the Royal Society of London* A133.821 (1931), pp. 60–72. ISSN: 0950-1207. DOI: 10.1098/rspa.1931.0130.

[107] S. G. Djorgovski et al. "Data-mining a large digital sky survey: from the challenges to the scientific results". In: *Proceedings of the International Society for Optics and Photonics* 3164 (1997), p. 98. DOI: 10.1117/12.292750. arXiv: astro-ph/9708218 [astro-ph].

[108] S. G. Djorgovski et al. "On the threshold of the reionization epoch". In: *The Astrophysical Journal* 560 (2001), p. L5. DOI: 10.1086/324175. arXiv: astro-ph/0108069 [astro-ph].

[109] S. Dodelson. *Modern cosmology*. San Diego, CA: Academic Press, 2003. URL: https://cds.cern.ch/record/1282338.

[110] S. Dodelson and L. M. Widrow. "Sterile-neutrinos as dark matter". In: *Physical Review Letters* 72 (1994), pp. 17–20. DOI: 10.1103/PhysRevLett.72.17. arXiv: hep-ph/9303287 [hep-ph].

[111] L. Dolan and R. Jackiw. "Symmetry behavior at finite temperature". In: *Physical Review* D9 (12 1974), pp. 3320–3341. DOI: 10.1103/PhysRevD.9.3320.

[112] J. N. Dossett et al. "Constraints and tensions in testing general relativity from Planck and CFHTLenS data including intrinsic alignment systematics". In: *Physical Review* D92.2 (2015), p. 023003. DOI: 10.1103/PhysRevD.92.023003. arXiv: 1501.03119 [astro-ph.CO].

[113] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2014.

[114] S. E. Dreyfus. "Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure". In: *Journal of Guidance, Control, and Dynamics* 13.5 (1990), pp. 926–928. ISSN: 0731-5090. DOI: 10.2514/3.25422.

[115] J-F. Dufaux, D. G. Figueroa, and J. Garcia-Bellido. "Gravitational Waves from Abelian Gauge Fields and Cosmic Strings at Preheating". In: *Physical Review* D82 (2010), p. 083518. DOI: 10.1103/PhysRevD.82.083518. arXiv: 1006.0217 [astro-ph.CO].

[116] R. Durrer and M. Kunz. "Cosmic microwave background anisotropies from scaling seeds: Generic properties of the correlation functions". In: *Physical Review* D57 (1998), R3199–R3203. DOI: 10.1103/PhysRevD.57.3199. arXiv: astro-ph/9711133 [astro-ph].

[117] G. R. Dvali, Q. Shafi, and R. K. Schaefer. "Large scale structure and supersymmetric inflation without fine tuning". In: *Physical Review Letters* 73 (1994), pp. 1886–1889. DOI: 10.1103/PhysRevLett.73.1886. arXiv: hep-ph/9406319 [hep-ph].

[118] G. R. Dvali and S. H. H. Tye. "Brane inflation". In: *Physics Letters* B450 (1999), pp. 72–82. DOI: 10.1016/S0370-2693(99)00132-X. arXiv: hep-ph/9812483 [hep-ph].

[119] G. R. Dvali and A. Vilenkin. "Formation and evolution of cosmic D strings". In: *Journal of Cosmology and Astroparticle Physics* 0403 (2004), p. 010. DOI: 10.1088/1475-7516/2004/03/010. arXiv: hep-th/0312007 [hep-th].

[120] C. Dvorkin, M. Wyman, and W. Hu. "Cosmic String constraints from WMAP and the South Pole Telescope". In: *Physical Review* D84 (2011), p. 123519. DOI: 10.1103/PhysRevD.84.123519. arXiv: 1109.4947 [astro-ph.CO].

[121] A. R. Edmonds. *Angular Momentum in Quantum Mechanics*. Investigations in Physics Series. Princeton University Press, 1996.

[122] G. Efstathiou, W. J. Sutherland, and S. J. Maddox. "The cosmological constant and cold dark matter". In: *Nature* 348 (Dec. 1990), pp. 705–707. DOI: 10.1038/348705a0.

[123] M. B. Einhorn, D. L. Stein, and D. Toussaint. "Are grand unified theories compatible with standard cosmology?" In: *Physical Review* D21 (12 1980), pp. 3295–3298. DOI: 10.1103/PhysRevD.21.3295.

[124] A. Einstein. "The Foundation of the General Theory of Relativity". In: *Annalen Phys.* 49 (1916), pp. 769–822.

[125] V. R. Eke et al. "Measuring Omega_0 using cluster evolution". In: *Monthly Notices of the Royal Astronomical Society* 298 (Aug. 1998), pp. 1145–1158. DOI: 10.1046/j.1365-8711.1998.01713.x. eprint: astro-ph/9802350.

[126] O. Elgaroy, M. Gramann, and O. Lahav. "Features in the primordial power spectrum: Constraints from the CMB and the limitation of the 2DF and SDSS redshift surveys to detect them". In: *Monthly Notices of the Royal Astronomical Society* 333 (2002), p. 93. DOI: 10.1046/j.1365-8711.2002.05382.x. arXiv: astro-ph/0111208 [astro-ph].

[127] O. Elgaroy and O. Lahav. "Neutrino masses from cosmological probes". In: *New Journal of Physics* 7 (2005), p. 61. DOI: 10.1088/1367-2630/7/1/061. arXiv: hep-ph/0412075 [hep-ph].

[128] J. L. Elman. "Finding structure in time". In: *Cognitive Science* 14.2 (1990), pp. 179–211. DOI: 10.1016/0364-0213(90)90002-E. URL: http://groups.lis.illinois.edu/amag/langev/paper/elman90findingStructure.html.

[129] A. van Engelen et al. "A Measurement of Gravitational Lensing of the Microwave Background Using South Pole Telescope Data". In: *The Astrophysical Journal* 756, 142 (Sept. 2012), p. 142. DOI: 10.1088/0004-637X/756/2/142. arXiv: 1202.0546 [astro-ph.CO].

[130] A. E. Everett, T. Vachaspati, and A. Vilenkin. "Monopole annihilation and causality". In: *Physical Review* D31 (8 1985), pp. 1925–1930. DOI: 10.1103/PhysRevD.31.1925.

[131] D. J. Fixsen. "The Temperature of the Cosmic Microwave Background". In: *The Astrophysical Journal* 707 (Dec. 2009), pp. 916–920. DOI: 10.1088/0004-637X/707/2/916. arXiv: 0911.1955.

[132] I. Florescu. *Probability and Stochastic Processes*. Wiley, 2014.

[133] K. Freese, M. S. Turner, and D. N. Schramm. "Monopole Catalysis of Nucleon Decay in Old Pulsars". In: *Physical Review Letters* 51 (18 1983), pp. 1625–1628. DOI: 10.1103/PhysRevLett.51.1625.

[134] J. H. Friedman. "Recent advances in predictive (machine) learning". In: *Journal of Classification* 23 (2006), pp. 175–197. DOI: 10.1007/s00357-006-0012-4.

[135] A. Friedmann. "Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes". In: *Zeitschrift für Physik* 21.1 (1924), pp. 326–332. ISSN: 0044-3328. DOI: 10.1007/BF01328280.

[136] G. Gamow. "The Evolution of the Universe". In: *Nature* 162 (Oct. 1948), pp. 680–682. DOI: 10.1038/162680a0.

[137] A. Gangui, L. Pogosian, and S. Winitzki. "CMB bispectrum from active models of structure formation". In: *Physical Review* D64 (2001), p. 043001. DOI: 10.1103/PhysRevD.64.043001. arXiv: astro-ph/0101453 [astro-ph].

[138] Q. Gao and Y. Gong. "The tension on the cosmological parameters from different observational data". In: *Classical and Quantum Gravity* 31 (2014), p. 105007. DOI: 10.1088/0264-9381/31/10/105007. arXiv: 1308.5627 [astro-ph.CO].

[139] W. A. Gardner. "Learning characteristics of stochastic-gradient-descent algorithms: A general study, analysis, and critique". In: *Signal Processing* 6.2 (1984), pp. 113 –133. ISSN: 0165-1684. DOI: 10.1016/0165-1684(84)90013-6.

[140] H. Georgi and S. L. Glashow. "Unity of all elementary-particle forces". In: *Physical Review Letters* 32.8 (1974), p. 438.

[141] F. A. Gers, J. Schmidhuber, and F. Cummins. "Learning to forget: Continual prediction with LSTM". In: *Neural computation* 12.10 (), p. 2451.

[142] H. Gil-Marín et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: BAO measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies". In: *Monthly Notices of the Royal Astronomical Society* 460.4 (2016), pp. 4210–4219. DOI: 10.1093/mnras/stw1264. arXiv: 1509.06373 [astro-ph.CO].

[143] H. Gil-Marín et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the LOS-dependent power spectrum of DR12 BOSS galaxies". In: *Monthly Notices of the Royal Astronomical Society* 460.4 (2016), pp. 4188–4209. DOI: 10.1093/mnras/stw1096. arXiv: 1509.06386 [astro-ph.CO].

[144] C. Giunti and M. Laveder. "Statistical Significance of the Gallium Anomaly". In: *Physical Review* C83 (2011), p. 065504. DOI: 10.1103/PhysRevC.83.065504. arXiv: 1006.3244 [hep-ph].

[145] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nuclear Physics* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[146] X. Glorot, A. Bordes, and Y. Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach". In: *Proceedings of the 28th international conference on machine learning (ICML-11)*. 2011, pp. 513–520.

[147] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[148] J. R. Gott III. "Gravitational lensing effects of vacuum strings: Exact solutions". In: *The Astrophysical Journal* 288 (1985), pp. 422–427. DOI: 10.1086/162808.

[149] M. Gramann and G. Hutsi. "Primordial feature at the scale of superclusters of galaxies". In: *Monthly Notices of the Royal Astronomical Society* 327 (2001), p. 538. DOI: 10.1046/j.1365-8711.2001.04733.x. arXiv: astro-ph/0102466 [astro-ph].

[150] L. M. Griffiths, J. Silk, and S. Zaroubi. "Bumpy power spectra and delta-T/T". In: *Monthly Notices of the Royal Astronomical Society* 324 (2001), p. 712. DOI: 10.1046/j.1365-8711.2001.04372.x. arXiv: astro-ph/0010571 [astro-ph].

[151] J. E. Gunn and B. A. Peterson. "On the Density of Neutral Hydrogen in Intergalactic Space." In: *The Astrophysical Journal* 142 (Nov. 1965), pp. 1633–1641. DOI: 10.1086/148444.

[152] A. H. Guth. "Inflationary universe: A possible solution to the horizon and flatness problems". In: *Physical Review* D23 (2 1981), pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

[153] A. H. Guth and S. H. H. Tye. "Phase Transitions and Magnetic Monopole Production in the Very Early Universe". In: *Physical Review Letters* 44 (10 1980), pp. 631–635. DOI: 10.1103/PhysRevLett.44.631.

[154] G. H, Z. Haiman, and J. J. Mohr. "Constraints on Ω_m, Ω_Λ, and σ_8 from Galaxy Cluster Redshift Distributions". In: *The Astrophysical Journal Letters* 560.2 (2001), p. L111. URL: http://stacks.iop.org/1538-4357/560/i=2/a=L111.

[155] A. C. Hall and A. Challinor. "Probing the neutrino mass hierarchy with CMB weak lensing". In: *Monthly Notices of the Royal Astronomical Society* 425 (2012), pp. 1170–1184. DOI: 10.1111/j.1365-2966.2012.21493.x. arXiv: 1205.6172 [astro-ph.CO].

[156] J. Hamann and J. Hasenkamp. "A new life for sterile neutrinos: resolving inconsistencies using hot dark matter". In: *Journal of Cosmology and Astroparticle Physics* 1310 (2013), p. 044. DOI: 10.1088/1475-7516/2013/10/044. arXiv: 1308.3255 [astro-ph.CO].

[157] A. Hanany and K. Hashimoto. "Reconnection of colliding cosmic strings". In: *Journal of High Energy Physics* 0506 (2005), p. 021. DOI: 10.1088/1126-6708/2005/06/021. arXiv: hep-th/0501031 [hep-th].

[158] S. Hannestad. "Neutrino physics from precision cosmology". In: *Progress in Particle and Nuclear Physics* 65 (2010), pp. 185–208. DOI: 10.1016/j.ppnp.2010.07.001. arXiv: 1007.0658 [hep-ph].

[159] S. Hannestad, S. H. Hansen, and F. L. Villante. "Probing the power spectrum bend with recent CMB data". In: *Astroparticle Physics* 16 (2001), pp. 137–144. DOI: 10.1016/S0927-6505(01)00103-7. arXiv: astro-ph/0012009 [astro-ph].

[160] E. R. Harrison. "Fluctuations at the Threshold of Classical Cosmology". In: *Physical Review* (May 1970), pp. 2726–2730. DOI: 10.1103/PhysRevD.1.2726.

[161] J. B. Hartle. *Gravity: An Introduction to Einstein's General Relativity*. Addison-Wesley, 2003.

[162] M. Hasselfield et al. "The Atacama Cosmology Telescope: Sunyaev-Zel'dovich selected galaxy clusters at 148 GHz from three seasons of data". In: *Journal of Cosmology and Astroparticle Physics* 1307 (2013), p. 008. DOI: 10.1088/1475-7516/2013/07/008. arXiv: 1301.0816 [astro-ph.CO].

[163] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109.

[164] D. M. Hawkins. "The Problem of Overfitting". In: *Journal of Chemical Information and Computer Sciences* 44.1 (2004). PMID: 14741005, pp. 1–12. DOI: 10.1021/ci0342472.

[165] D. K. Hazra, A. Shafieloo, and G. F. Smoot. "Reconstruction of broad features in the primordial spectrum and inflaton potential from Planck". In: *Journal of Cosmology and Astroparticle Physics* 1312 (2013), p. 035. DOI: 10.1088/1475-7516/2013/12/035. arXiv: 1310.3038 [astro-ph.CO].

[166] K. He et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *ArXiv e-prints* (Feb. 2015). arXiv: 1502.01852 [cs.CV].

[167] A. Heavens et al. "Bayesian hierarchical modelling of weak lensing - the golden goal". In: 2016. arXiv: 1602.05345 [astro-ph.CO].

[168] J. P. Henry. "A Measurement of the Density Parameter Derived from the Evolution of Cluster X-Ray Temperatures". In: *The Astrophysical Journal Letters* 489 (Nov. 1997), p. L1. DOI: 10.1086/310949.

[169] C. Heymans et al. "CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey". In: *Monthly Notices of the Royal Astronomical Society* 427 (Nov. 2012), pp. 146–166. DOI: 10.1111/j.1365-2966.2012.21952.x. arXiv: 1210.0032 [astro-ph.CO].

[170] C. Heymans et al. "CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments". In: *Monthly Notices of the Royal Astronomical Society* 432 (2013), p. 2433. DOI: 10.1093/mnras/stt601. arXiv: 1303.1808 [astro-ph.CO].

[171] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[172] M. Hindmarsh. "Semilocal topological defects". In: *Nuclear Physics* B392 (1993), pp. 461–492. DOI: 10.1016/0550-3213(93)90681-E. arXiv: hep-ph/9206229 [hep-ph].

[173] M. Hindmarsh and P. M. Saffin. "Scaling in a SU(2)/Z_3 model of cosmic superstring networks". In: *Journal of High Energy Physics* 08 (2006), p. 066. DOI: 10.1088/1126-6708/2006/08/066. arXiv: hep-th/0605014 [hep-th].

[174] M. Hindmarsh, S. Stuckey, and N. Bevis. "Abelian Higgs Cosmic Strings: Small Scale Structure and Loops". In: *Physical Review* D79 (2009), p. 123504. DOI: 10.1103/PhysRevD.79.123504. arXiv: 0812.1929 [hep-th].

[175] M. B. Hindmarsh and T. W. B. Kibble. "Cosmic strings". In: *Reports on Progress in Physics* 58 (1995), pp. 477–562. DOI: 10.1088/0034-4885/58/5/001. arXiv: hep-ph/9411342 [hep-ph].

[176] G. Hinshaw et al. "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results". In: *The Astrophysical Journal Supplement* 208, 19 (Oct. 2013), p. 19. DOI: 10.1088/0067-0049/208/2/19. arXiv: 1212.5226 [astro-ph.CO].

[177] G. Hinshaw et al. "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results". In: *The Astrophysical Journal Supplement* 208 (2013), p. 19. DOI: 10.1088/0067-0049/208/2/19. arXiv: 1212.5226 [astro-ph.CO].

[178] M. P. Hobson et al. *Bayesian Methods in Cosmology*. Cambridge University Press, 2014.

[179] S. Hochreiter and J. Schmidhuber. "Long short-term memory". In: *Neural Computation* 9(8) (1997), pp. 1735–1780.

[180] J. Hong, J. Kim, and P. Sikivie. "Wiggly relativistic strings". In: *Physical Review Letters* 69 (18 1992), pp. 2611–2614. DOI: 10.1103/PhysRevLett.69.2611.

[181] G.'t Hooft. "Magnetic monopoles in unified gauge theories". In: *Nuclear Physics* B79.2 (1974), pp. 276 –284. ISSN: 0550-3213. DOI: 10.1016/0550-3213(74)90486-6.

[182] C. Howlett et al. "CMB power spectrum parameter degeneracies in the era of precision cosmology". In: *Journal of Cosmology and Astroparticle Physics* 1204 (2012), p. 027. DOI: 10.1088/1475-7516/2012/04/027. arXiv: 1201.3654 [astro-ph.CO].

[183] B. Hu and M. Raveri. "Can modified gravity models reconcile the tension between the CMB anisotropy and lensing maps in Planck-like observations?" In: *Physical Review* D91.12 (2015), p. 123515. DOI: 10.1103/PhysRevD.91.123515. arXiv: 1502.06599 [astro-ph.CO].

[184] W. Hu. "Wandering in the Background: A CMB Explorer". PhD thesis. UC, Berkeley, 1995. arXiv: astro-ph/9508126 [astro-ph]. URL: http://alice.cern.ch/format/showfull?sysnb=0207836.

[185] W. Hu. "Weak lensing of the CMB: A harmonic approach". In: *Physical Review* D62 (2000), p. 043007. DOI: 10.1103/PhysRevD.62.043007. arXiv: astro-ph/0001303 [astro-ph].

[186] W. Hu, D. J. Eisenstein, and M. Tegmark. "Weighing neutrinos with galaxy surveys". In: *Physical Review Letters* 80 (1998), pp. 5255–5258. DOI: 10.1103/PhysRevLett.80.5255. arXiv: astro-ph/9712057.

[187] W. Hu and M. J. White. "Acoustic signatures in the cosmic microwave background". In: *The Astrophysical Journal* 471 (1996), pp. 30–51. DOI: 10.1086/177951. arXiv: astro-ph/9602019 [astro-ph].

[188] E. Hubble. "A relation between distance and radial velocity among extra-galactic nebulae". In: *Proceedings of the National Academy of Sciences* 15.3 (1929), pp. 168–173. DOI: 10.1073/pnas.15.3.168.

[189] D. Huterer and T. Vachaspati. "Gravitational lensing by cosmic strings in the era of wide - field surveys". In: *Physical Review* D68 (2003), p. 041301. DOI: 10.1103/PhysRevD.68.041301. arXiv: astro-ph/0305006 [astro-ph].

[190] H. F. Inman and E. L. Bradley Jr. "The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities". In: *Communications in Statistics - Theory and Methods* 18.10 (1989), pp. 3851–3874. DOI: 10.1080/03610928908830127.

[191] J. C. Jackson. "A Critique of Rees's Theory of Primordial Gravitational Radiation". In: *Monthly Notices of the Royal Astronomical Society* 156.1 (1972), 1P. DOI: 10.1093/mnras/156.1.1P.

[192] M. G. Jackson. "Interactions of cosmic superstrings". In: *Journal of High Energy Physics* 0709 (2007), p. 035. DOI: 10.1088/1126-6708/2007/09/035. arXiv: 0706.1264 [hep-th].

[193] M. G. Jackson, N. T. Jones, and J. Polchinski. "Collisions of cosmic F and D-strings". In: *Journal of High Energy Physics* 0510 (2005), p. 013. DOI: 10.1088/1126-6708/2005/10/013. arXiv: hep-th/0405229 [hep-th].

[194] R. Jeannerot, J. Rocher, and M. Sakellariadou. "How generic is cosmic string formation in SUSY GUTs". In: *Physical Review* D68 (2003), p. 103514. DOI: 10.1103/PhysRevD.68.103514. arXiv: hep-ph/0308134 [hep-ph].

[195] Sir H. Jeffreys. *Scientific Inference*. 3rd ed. Cambridge, UK: Cambridge University Press, 1973, p. 31.

[196] Sir H. Jeffreys. *The Theory of Probability*. New York, NY, USA: Oxford University Press, 1961, p. 432.

[197] D. H. Jones et al. "The 6dF Galaxy Survey: Samples, observational techniques and the first data release". In: *Monthly Notices of the Royal Astronomical Society* 355 (2004), pp. 747–763. DOI: 10.1111/j.1365-2966.2004.08353.x. arXiv: astro-ph/0403501 [astro-ph].

[198] N. T. Jones, H. Stoica, and S. H. H. Tye. "The Production, spectrum and evolution of cosmic strings in brane inflation". In: *Physics Letters* B563 (2003), pp. 6–14. DOI: 10.1016/S0370-2693(03)00592-6. arXiv: hep-th/0303269 [hep-th].

[199] S. Joudaki et al. "CFHTLenS revisited: assessing concordance with Planck including astrophysical systematics". In: *Monthly Notices of the Royal Astronomical Society* 465.2 (2017), pp. 2033–2052–2052. DOI: 10.1093/mnras/stw2665. arXiv: 1601.05786 [astro-ph.CO].

[200] S. Kachru et al. "Towards inflation in string theory". In: *Journal of Cosmology and Astroparticle Physics* 0310 (2003), p. 013. DOI: 10.1088/1475-7516/2003/10/013. arXiv: hep-th/0308055 [hep-th].

[201] D. I. Kaiser, E. A. Mazenc, and E. I. Sfakianakis. "Primordial Bispectrum from Multifield Inflation with Nonminimal Couplings". In: *Physical Review* D87.6 (2013), p. 064004. DOI: 10.1103/PhysRevD.87.064004. arXiv: 1210.7487 [astro-ph.CO].

[202] N. Kaiser. "Clustering in real space and in redshift space". In: *Monthly Notices of the Royal Astronomical Society* 227 (July 1987), pp. 1–21. DOI: 10.1093/mnras/227.1.1.

[203] N. Kaiser and A. Stebbins. "Microwave Anisotropy Due to Cosmic Strings". In: *Nature* 310 (1984), pp. 391–393. DOI: 10.1038/310391a0.

[204] N. Kaiser et al. "Mapping the dark matter in clusters". In: (1994). arXiv: astro-ph/9407004 [astro-ph].

[205] M. Kamionkowski, A. Kosowsky, and A. Stebbins. "A Probe of primordial gravity waves and vorticity". In: *Physical Review Letters* 78 (1997), pp. 2058–2061. DOI: 10.1103/PhysRevLett.78.2058. arXiv: astro-ph/9609132 [astro-ph].

[206] M. Kamionkowski, A. Kosowsky, and A. Stebbins. "Statistics of cosmic microwave background polarization". In: *Physical Review* D55 (1997), pp. 7368–7388. DOI: 10.1103/PhysRevD.55.7368. arXiv: astro-ph/9611125 [astro-ph].

[207] M. Kaplinghat et al. "Probing the reionization history of the universe using the cosmic microwave background polarization". In: *The Astrophysical Journal* 583 (2003), pp. 24–32. DOI: 10.1086/344927. arXiv: astro-ph/0207591 [astro-ph].

[208] N. V. Karpenka, F. Feroz, and M. P. Hobson. "A simple and robust method for automated photometric classification of supernovae using neural networks". In: *Monthly Notices of the Royal Astronomical Society* 429 (2013), p. 1278. DOI: 10.1093/mnras/sts412. arXiv: 1208.1264 [astro-ph.CO].

[209] T. Katori. "Short Baseline Neutrino Oscillation Experiments". In: *Journal of Physics: Conference Series* 598.1 (2015), p. 012006. DOI: 10.1088/1742-6596/598/1/012006. arXiv: 1404.6882 [hep-ph].

[210] M. Kawasaki, K. Miyamoto, and K. Nakayama. "Gravitational waves from kinks on infinite cosmic strings". In: *Physical Review* D81 (2010), p. 103523. DOI: 10.1103/PhysRevD.81.103523. arXiv: 1002.0652 [astro-ph.CO].

[211] R. Kennedy et al. "Galaxy And Mass Assembly (GAMA): understanding the wavelength dependence of galaxy structure with bulge-disc decompositions". In: *Monthly Notices of the Royal Astronomical Society* 460 (Aug. 2016), pp. 3458–3471. DOI: 10.1093/mnras/stw1176. arXiv: 1605.04720.

[212] R. Kessler et al. "Results from the Supernova Photometric Classification Challenge". In: *Publications of the Astronomical Society of the Pacific* 122 (2010), pp. 1415–1431. DOI: 10.1086/657607. arXiv: 1008.1024 [astro-ph.CO].

[213] R. Kessler et al. "Supernova Photometric Classification Challenge". In: (2010). arXiv: 1001.5210 [astro-ph.IM].

[214] T. W. B. Kibble. "Evolution of a system of cosmic strings". In: *Nuclear Physics* B252 (1985), p. 227. DOI: 10.1016/0550-3213(85)90439-0.

[215] T. W. B. Kibble. "Some Implications of a Cosmological Phase Transition". In: *Physics Reports* 67 (1980), p. 183. DOI: 10.1016/0370-1573(80)90091-5.

[216] T. W. B. Kibble. "Topology of cosmic domains and strings". In: *Journal of Physics* A9.8 (1976), p. 1387. URL: http://stacks.iop.org/0305-4470/9/i=8/a=029.

[217] T. W. B. Kibble and N. G. Turok. "Cosmic Strings and Galaxy Formation". In: *Philosophical Transactions of the Royal Society of London* A320.1556 (1986), pp. 565–571. ISSN: 0080-4614. DOI: 10.1098/rsta.1986.0137.

[218] M. Kilbinger. "Cosmology with cosmic shear observations: a review". In: *Reports on Progress in Physics* 78 (2015), p. 086901. DOI: 10.1088/0034-4885/78/8/086901. arXiv: 1411.0115 [astro-ph.CO].

[219] M. Kilbinger et al. "Precision calculations of the cosmic shear power spectrum projection". In: (2017). arXiv: 1702.05301 [astro-ph.CO].

[220] L. J. King. "Cosmic shear as a tool for precision cosmology: minimising intrinsic galaxy alignment-lensing interference". In: *Astronomy & Astrophysics* 441.1 (2005), pp. 47–53. DOI: 10.1051/0004-6361:20053330.

[221] D. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: *ArXiv e-prints* (Dec. 2014). arXiv: 1412.6980 [cs.LG].

[222] T. D. Kitching et al. "Discrepancies between CFHTLenS cosmic shear and Planck: new physics or systematic effects?" In: *Monthly Notices of the Royal Astronomical Society* 459.1 (2016), pp. 971–981. DOI: 10.1093/mnras/stw707. arXiv: 1602.02960 [astro-ph.CO].

[223] A. Kogut et al. "Wilkinson Microwave Anisotropy Probe (WMAP) first year observations: TE polarization". In: *The Astrophysical Journal Supplement* 148 (2003), p. 161. DOI: 10.1086/377219. arXiv: astro-ph/0302213 [astro-ph].

[224] E. W. Kolb, S. A. Colgate, and Jeffrey A. Harvey. "Monopole Catalysis of Nucleon Decay in Neutron Stars". In: *Physical Review Letters* 49 (1982), p. 1373. DOI: 10.1103/PhysRevLett.49.1373.

[225] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in Neural Information Processing Systems*. 2012.

[226] K. Kuijken et al. "Gravitational Lensing Analysis of the Kilo Degree Survey". In: *Monthly Notices of the Royal Astronomical Society* 454.4 (2015), pp. 3500–3532. DOI: 10.1093/mnras/stv2140. arXiv: 1507.00738 [astro-ph.CO].

[227] A. Lazanu, E. P. S. Shellard, and M. Landriau. "CMB power spectrum of Nambu-Goto cosmic strings". In: *Physical Review* D91.8 (2015), p. 083519. DOI: 10.1103/PhysRevD.91.083519. arXiv: 1410.4860 [astro-ph.CO].

[228] Y. Lecun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521.7553 (May 2015), pp. 436–444. ISSN: 0028-0836. DOI: 10.1038/nature14539.

[229] Y. LeCun et al. "Efficient BackProp". In: *Neural Networks: Tricks of the Trade*. Ed. by G. B. Orr and K-R. Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 9–50. DOI: 10.1007/3-540-49430-8_2.

[230] B. Leistedt, H. V. Peiris, and L. Verde. "No new cosmological concordance with massive sterile neutrinos". In: *Physical Review Letters* 113 (2014), p. 041301. DOI: 10.1103/PhysRevLett.113.041301. arXiv: 1404.5950 [astro-ph.CO].

[231] G. Lemaître. "Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques". In: *Annales de la Société Scientifique de Bruxelles* 47.A (1927), pp. 49–59.

[232] J. Lesgourges et al. *Neutrino Cosmology*. Cambridge: Cambridge University Press, 2013.

[233] J. Lesgourgues and S. Pastor. "Massive neutrinos and cosmology". In: *Physics Reports* 429 (2006), pp. 307–379. DOI: 10.1016/j.physrep.2006.04.001. arXiv: astro-ph/0603494 [astro-ph].

[234] A. Lewis and S. Bridle. "Cosmological parameters from CMB and other data: A Monte Carlo approach". In: *Physical Review* D66 (2002), p. 103511. DOI: 10.1103/PhysRevD.66.103511. arXiv: astro-ph/0205436.

[235] A. Lewis and A. Challinor. "Evolution of cosmological dark matter perturbations". In: *Physical Review* D66 (2002), p. 023531. DOI: 10.1103/PhysRevD.66.023531. arXiv: astro-ph/0203507 [astro-ph].

[236] A. Lewis and A. Challinor. "Weak gravitational lensing of the CMB". In: *Physics Reports* 429 (2006), pp. 1–65. DOI: 10.1016/j.physrep.2006.03.002. arXiv: astro-ph/0601594 [astro-ph].

[237] A. Lewis, A. Challinor, and A. Lasenby. "Efficient computation of CMB anisotropies in closed FRW models". In: *The Astrophysical Journal* 538 (2000), pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

[238] C. J. Li and L. Yan. "Mechanical system modelling using recurrent neural networks via quasi-Newton learning methods". In: *Applied Mathematical Modelling* 19.7 (1995), pp. 421 –428. ISSN: 0307-904X. DOI: 10.1016/0307-904X(95)00015-C.

[239] A. R. Liddle. "An Introduction to cosmological inflation". In: *Proceedings, Summer School in High-energy physics and cosmology*. 1999, pp. 260–295. arXiv: astro-ph/9901124 [astro-ph]. URL: http://alice.cern.ch/format/showfull?sysnb=0301651.

[240] A. D. Linde. "Hybrid inflation". In: *Physical Review* D49 (1994), pp. 748–754. DOI: 10.1103/PhysRevD.49.748. arXiv: astro-ph/9307002 [astro-ph].

[241] J. Lizarraga et al. "Constraining topological defects with temperature and polarization anisotropies". In: *Physical Review* D90.10 (2014), p. 103504. DOI: 10.1103/PhysRevD.90.103504. arXiv: 1408.4126 [astro-ph.CO].

[242] M. Lochner et al. "Photometric Supernova Classification With Machine Learning". In: *The Astrophysical Journal Supplement* 225.2 (2016), p. 31. DOI: 10.3847/0067-0049/225/2/31. arXiv: 1603.00882 [astro-ph.IM].

[243] C-P. Ma and E. Bertschinger. "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges". In: *The Astrophysical Journal* 455 (1995), pp. 7–25. DOI: 10.1086/176550. arXiv: astro-ph/9506072 [astro-ph].

[244] A. L. Maas, A. Y. Hannun, and A. Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Proceedings of the International Machine Learning Society*. Vol. 30. 1. 2013.

[245] N. MacCrann et al. "Cosmic Discordance: Are Planck CMB and CFHTLenS weak lensing measurements out of tune?" In: *Monthly Notices of the Royal Astronomical Society* 451.3 (2015), pp. 2877–2888. DOI: 10.1093/mnras/stv1154. arXiv: 1408.4742 [astro-ph.CO].

[246] A. B. Mantz et al. "Weighing the giants – IV. Cosmology and neutrino mass". In: *Monthly Notices of the Royal Astronomical Society* 446 (2015), pp. 2205–2225. DOI: 10.1093/mnras/stu2096. arXiv: 1407.4516 [astro-ph.CO].

[247] P. Marshall, N. Rajguru, and A. Slosar. "Bayesian evidence as a tool for comparing datasets". In: *Physical Review* D73 (2006), p. 067302. DOI: 10.1103/PhysRevD.73.067302. arXiv: astro-ph/0412535 [astro-ph].

[248] C. J. A. P. Martins and E. P. S. Shellard. "Extending the velocity dependent one scale string evolution model". In: *Physical Review* D65 (2002), p. 043514. DOI: 10.1103/PhysRevD.65.043514. arXiv: hep-ph/0003298 [hep-ph].

[249] C. J. A. P. Martins and E. P. S. Shellard. "Fractal properties and small-scale structure of cosmic string networks". In: *Physical Review* D73 (2006), p. 043515. DOI: 10.1103/PhysRevD.73.043515. arXiv: astro-ph/0511792 [astro-ph].

[250] C. J. A. P. Martins and E. P. S. Shellard. "Quantitative string evolution". In: *Physical Review* D54 (1996), pp. 2535–2556. DOI: 10.1103/PhysRevD.54.2535. arXiv: hep-ph/9602271 [hep-ph].

[251] C. J. A. P. Martins and E. P. S. Shellard. "String evolution with friction". In: *Physical Review* D53 (1996), pp. 575–579. DOI: 10.1103/PhysRevD.53.575. arXiv: hep-ph/9507335 [hep-ph].

[252] C. J. A. P. Martins, E. P. S. Shellard, and J. P. P. Vieira. "Models for Small-Scale Structure on Cosmic Strings: Mathematical Formalism". In: *Physical Review* D90.4 (2014), p. 043518. DOI: 10.1103/PhysRevD.90.043518. arXiv: 1405.7722 [hep-ph].

[253] J. C. Mather et al. "Calibrator design for the COBE far infrared absolute spectrophotometer (FIRAS)". In: *The Astrophysical Journal* 512 (1999), pp. 511–520. DOI: 10.1086/306805. arXiv: astro-ph/9810373 [astro-ph].

[254] J. C. Mather et al. "Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument". In: *The Astrophysical Journal* 420 (Jan. 1994), pp. 439–444. DOI: 10.1086/173574.

[255] T. Matheson et al. "Spectroscopy in the Era of LSST". In: *ArXiv e-prints* (Nov. 2013). arXiv: 1311.2496 [astro-ph.CO].

[256] J McCaffrey. *Test Run - L1 and L2 Regularization for Machine Learning*. https://msdn.microsoft.com/en-us/magazine/dn904675.aspx?ranMID=24542&ranEAID=TnL5HPStwNw&ranSiteID=TnL5HPStwNw-_DETSdukDsUTyKLFf22XiA&tduid=(c50775b88aaf03ce946f8c54b74d64e5)(256380)(2459594)(TnL5HPStwNw-_DETSdukDsUTyKLFf22XiA)(). Blog. 2015.

[257] W. S. McCulloch and W. Pitts. "A logical calculus of the ideas immanent in nervous activity". In: *The Bulletin of Mathematical Biophysics* 5.4 (1943), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259.

[258] P. McDonald and A. Roy. "Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS". In: *Journal of Cosmology and Astroparticle Physics* 0908 (2009), p. 020. DOI: 10.1088/1475-7516/2009/08/020. arXiv: 0902.0991 [astro-ph.CO].

[259] W. D. McGlinn. *Introduction to Relativity*. Introduction to Relativity. Johns Hopkins University Press, 2003.

[260] L. Medsker. *Recurrent Neural Networks: Design and Applications*. International Series on Computational Intelligence. CRC-Press, 2000.

[261] G. Mention et al. "The Reactor Antineutrino Anomaly". In: *Physical Review* D83 (2011), p. 073006. DOI: 10.1103/PhysRevD.83.073006. arXiv: 1101.2755 [hep-ex].

[262] N. Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *Journal of Computational Physics* 21 (June 1953), pp. 1087–1092. DOI: 10.1063/1.1699114.

[263] A. Mirizzi et al. "The strongest bounds on active-sterile neutrino mixing after Planck data". In: *Physics Letters* B726 (2013), pp. 8–14. DOI: 10.1016/j.physletb.2013.08.015. arXiv: 1303.5368 [astro-ph.CO].

[264] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Gravitation. Freeman, W. H., 1973.

[265] V. F. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, 2005.

[266] P. Mukherjee et al. "Detecting and distinguishing topological defects in future data from the CMBPol satellite". In: *Physical Review* D83 (2011), p. 043003. DOI: 10.1103/PhysRevD.83.043003. arXiv: 1010.5662 [astro-ph.CO].

[267] V. Nair and G. E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010, pp. 807–814.

[268] Y. Nambu. *International Conference of Symmetries and Quark Models: Proceedings*. Chand, R.: Wayne State University, 1969.

[269] J. Newling et al. "Statistical Classification Techniques for Photometric Supernova Typing". In: *Monthly Notices of the Royal Astronomical Society* 414 (2011), pp. 1987–2004. DOI: 10.1111/j.1365-2966.2011.18514.x. arXiv: 1010.1005 [astro-ph.CO].

[270] H. B. Nielsen and P. Olesen. "Vortex-line models for dual strings". In: *Nuclear Physics B* 61 (1973), pp. 45 –61. ISSN: 0550-3213. DOI: 10.1016/0550-3213(73)90350-7.

[271] M. A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL: http://neuralnetworksanddeeplearning.com.

[272] C. Olah. *Understanding LSTM Networks*. http://colah.github.io/posts/2015-08-Understanding-LSTMs/. Blog. 2015.

[273] S. Olmez, V. Mandic, and X. Siemens. "Gravitational-Wave Stochastic Background from Kinks and Cusps on Cosmic Strings". In: *Physical Review* D81 (2010), p. 104028. DOI: 10.1103/PhysRevD.81.104028. arXiv: 1004.0890 [astro-ph.CO].

[274] C. A. J. O'Hare. "Dark matter astrophysical uncertainties and the neutrino floor". In: *Physical Review* D94.6 (2016), p. 063527. DOI: 10.1103/PhysRevD.94.063527. arXiv: 1604.03858 [astro-ph.CO].

[275] N. Padmanabhan et al. "A 2 per cent distance to z=0.35 by reconstructing baryon acoustic oscillations - I. Methods and application to the Sloan Digital Sky Survey". In: *Monthly Notices of the Royal Astronomical Society* 427.3 (2012), pp. 2132–2145. DOI: 10.1111/j.1365-2966.2012.21888.x. arXiv: 1202.0090 [astro-ph.CO].

[276] T. Padmanabhan. "Inverse Compton scattering – revisited". In: *Journal of Astrophysics and Astronomy* 18.1 (1997), pp. 87–90. ISSN: 0973-7758. DOI: 10.1007/BF02714856.

[277] E. N. Parker. "The Origin of Magnetic Fields". In: *The Astrophysical Journal* 160 (May 1970), p. 383. DOI: 10.1086/150442.

[278] J. C. Pati and A. Salam. "Lepton number as the fourth "color"". In: *Physical Review* D10 (1 1974), pp. 275–289. DOI: 10.1103/PhysRevD.10.275.

[279] B. A. Pearlmutter. "Learning state space trajectories in recurrent neural networks". In: *Neural Computation* 1 (2 1989), 263–269.

[280] P. J. E. Peebles. "Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations". In: *The Astrophysical Journal Letters* 263 (Dec. 1982), pp. L1–L5. DOI: 10.1086/183911.

[281] P. J. E. Peebles. "Recombination of the Primeval Plasma". In: *The Astrophysical Journal* 153 (July 1968), p. 1. DOI: 10.1086/149628.

[282] P. J. E. Peebles and J. T. Yu. "Primeval Adiabatic Perturbation in an Expanding Universe". In: *The Astrophysical Journal* 162 (Dec. 1970), p. 815. DOI: 10.1086/150713.

[283] U-L. Pen, U. Seljak, and N. Turok. "Power spectra in global defect theories of cosmic structure formation". In: *Physical Review Letters* 79 (1997), pp. 1611–1614. DOI: 10.1103/PhysRevLett.79.1611. arXiv: astro-ph/9704165 [astro-ph].

[284] A. A. Penzias and R. W. Wilson. "A Measurement of Excess Antenna Temperature at 4080 Mc/s." In: *The Astrophysical Journal* 142 (July 1965), pp. 419–421. DOI: 10.1086/148307.

[285] S. Perlmutter et al. "Measurements of Omega and Lambda from 42 High-Redshift Supernovae". In: *The Astrophysical Journal* 517 (June 1999), pp. 565–586. DOI: 10.1086/307221. eprint: astro-ph/9812133.

[286] W. Pitts and W. S. McCulloch. "How we know universals the perception of auditory and visual forms". In: *The Bulletin of Mathematical Biophysics* 9.3 (1947), pp. 127–147. ISSN: 1522-9602. DOI: 10.1007/BF02478291.

[287] L. Pogosian. *An Update On CMBACT*. Cosmic Strings 2014: Arizona State University, 2014.

[288] L. Pogosian and T. Vachaspati. "Cosmic microwave background anisotropy from wiggly strings". In: *Physical Review* D60 (1999), p. 083504. DOI: 10.1103/PhysRevD.60.083504. arXiv: astro-ph/9903361 [astro-ph].

[289] L. Pogosian and M. Wyman. "B-modes from cosmic strings". In: *Physical Review* D77 (2008), p. 083509. DOI: 10.1103/PhysRevD.77.083509. arXiv: 0711.0747 [astro-ph].

[290] L. Pogosian et al. "Observational constraints on cosmic string production during brane inflation". In: *Physical Review* D68 (2003), p. 023506. DOI: 10.1103/PhysRevD.68.023506, 10.1103/PhysRevD.73.089904, 10.1103/PhysRevD.68.02350610.1103/PhysRevD.73.089904. arXiv: hep-th/0304188 [hep-th].

[291] J. Polchinski. "Introduction to cosmic F- and D-strings". In: *String theory: From gauge interactions to cosmology. Proceedings, NATO Advanced Study Institute, Cargese, France, June 7-19, 2004*. 2004, pp. 229–253. arXiv: hep-th/0412244 [hep-th].

[292] A. M. Polyakov. "Particle Spectrum in the Quantum Field Theory". In: *Journal of Experimental and Theoretical Physics Letters* 20 (1974). [Pisma Zh. Eksp. Teor. Fiz.20,430(1974)], pp. 194–195.

[293] A. Pourtsidou et al. "Scaling configurations of cosmic superstring networks and cosmological implications". In: *Physical Review* D83 (2011), p. 063525. DOI: 10.1103/PhysRevD.83.063525. arXiv: 1012.5014 [astro-ph.CO].

[294] W. H. Press and P. Schechter. "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: *The Astrophysical Journal* 187 (Feb. 1974), pp. 425–438. DOI: 10.1086/152650.

[295] M. J. Rees and D. W. Sciama. "Large-scale Density Inhomogeneities in the Universe". In: *Nature* 217 (Feb. 1968), pp. 511–516. DOI: 10.1038/217511a0.

[296] D. Regan and M. Hindmarsh. "The bispectrum of cosmic string temperature fluctuations including recombination effects". In: *Journal of Cosmology and Astroparticle Physics* 1510.10 (2015), p. 030. DOI: 10.1088/1475-7516/2015/10/030. arXiv: 1508.02231 [astro-ph.CO].

[297] D. Regan and M. Hindmarsh. "The bispectrum of matter perturbations from cosmic strings". In: *Journal of Cosmology and Astroparticle Physics* 1503.03 (2015), p. 008. DOI: 10.1088/1475-7516/2015/03/008. arXiv: 1411.2641 [astro-ph.CO].

[298] C. L. Reichardt et al. "A Measurement of Secondary Cosmic Microwave Background Anisotropies with Two Years of South Pole Telescope Observations". In: *The Astrophysical Journal* 755, 70 (Aug. 2012), p. 70. DOI: 10.1088/0004-637X/755/1/70. arXiv: 1111.0932 [astro-ph.CO].

[299] J. A. Rice. *Mathematical Statistics and Data Analysis*. Advanced series. Cengage Learning, 2006.

[300] S. Riemer-Sørensen, D. Parkinson, and T. M. Davis. "Combining Planck data with large scale structure information gives a strong neutrino mass constraint". In: *Physical Review* D89 (2014), p. 103505. DOI: 10.1103/PhysRevD.89.103505. arXiv: 1306.4153 [astro-ph.CO].

[301] A. G. Riess et al. "A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3". In: *The Astrophysical Journal* 730 (2011), p. 119. DOI: 10.1088/0004-637X/732/2/129,10.1088/0004-637X/730/2/119. arXiv: 1103.2976 [astro-ph.CO].

[302] A. G. Riess et al. "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *The Astronomical Journal* 116 (Sept. 1998), pp. 1009–1038. DOI: 10.1086/300499. eprint: astro-ph/9805201.

[303] M. Ritthaler et al. "Bayesian Belief Networks for Astronomical Object Recognition and Classification in CTI-II". In: *Astronomical Data Analysis Software and Systems XVI*. Ed. by R. A. Shaw, F. Hill, and D. J. Bell. Vol. 376. Astronomical Society of the Pacific Conference Series. Oct. 2007, p. 413.

[304] G. O. Roberts, A. Gelman, and W. R. Gilks. "Weak convergence and optimal scaling of random walk Metropolis algorithms". In: *Annals of Applied Probability* 7.1 (Feb. 1997), pp. 110–120. DOI: 10.1214/aoap/1034625254.

[305] D. J. Rohde et al. "Applying machine learning to catalogue matching in astrophysics". In: *Monthly Notices of the Royal Astronomical Society* 360 (2005), pp. 69–75. DOI: 10.1111/j.1365-2966.2005.08930.x. arXiv: astro-ph/0504013 [astro-ph].

[306] L. Rosasco et al. "Are Loss Functions All the Same?" In: *Neural Computation* 16.5 (2004), pp. 1063–1076. ISSN: 0899-7667. DOI: 10.1162/089976604773135104.

[307] F. Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Report (Cornell Aeronautical Laboratory). Spartan Books, 1962.

[308] E. Rozo et al. "Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog". In: *The Astrophysical Journal* 708 (Jan. 2010), pp. 645–660. DOI: 10.1088/0004-637X/708/1/645. arXiv: 0902.3702 [astro-ph.CO].

[309] M. R. Rudary. *On Predictive Linear Gaussian Models*. ProQuest, 2009.

[310] E. J. Ruiz and D. Huterer. "Testing the dark energy consistency with geometry and growth". In: *Physical Review* D91 (2015), p. 063009. DOI: 10.1103/PhysRevD.91.063009. arXiv: 1410.5832 [astro-ph.CO].

[311] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536.

[312] R. K. Sachs and A. M. Wolfe. "Perturbations of a Cosmological Model and Angular Variations of the Microwave Background". In: *The Astrophysical Journal* 147 (Jan. 1967), p. 73. DOI: 10.1086/148982.

[313] S. Saito et al. "Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum". In: *Physical Review* D90.12 (2014), p. 123522. DOI: 10.1103/PhysRevD.90.123522. arXiv: 1405.1447 [astro-ph.CO].

[314] M. Sako et al. "The Sloan Digital Sky Survey-II Supernova Survey: Search Algorithm and Follow-up Observations". In: *The Astronomical Journal* 135, 348-373 (Jan. 2008), pp. 348–373. DOI: 10.1088/0004-6256/135/1/348. arXiv: 0708.2750.

[315] A. Salam. "Gauge Unification of Fundamental Forces". In: *Reviews of Modern Physics* 52 (1980). [Science210,723(1980)], pp. 525–538. DOI: 10.1103/RevModPhys.52.525.

[316] N. G. Sànchez and Y. N. Parijskij. *The Early Universe and the Cosmic Microwave Background: Theory and Observations*. Nato Science Series II. Springer Netherlands, 2012.

[317] A. Sandage. "Photoelectric Observations of the Interacting Galaxies VV 117 and VV 123 Related to the Time of Formation of Their Satellites." In: *The Astrophysical Journal* 138 (Oct. 1963), p. 863. DOI: 10.1086/147691.

[318] S. Sarangi and S. H. H. Tye. "Cosmic string production towards the end of brane inflation". In: *Physics Letters* B536 (2002), pp. 185–192. DOI: 10.1016/S0370-2693(02)01824-5. arXiv: hep-th/0204074 [hep-th].

[319] A. Saro et al. "Constraints on the CMB temperature evolution using multiband measurements of the Sunyaev–Zel'dovich effect with the South Pole Telescope". In: *Monthly Notices of the Royal Astronomical Society* 440.3 (2014), pp. 2610–2615. DOI: 10.1093/mnras/stu575. arXiv: 1312.2462 [astro-ph.CO].

[320] J. Schmidhuber. "A local learning algorithm for dynamic feedforward and recurrent networks". In: *Connection Science* 1 (4 1989), 403–412.

[321] J. Schmidhuber. "Deep Learning in Neural Networks: An Overview". In: *ArXiv e-prints* (Apr. 2014). arXiv: 1404.7828.

[322] J. H. Schwarz. "An SL(2,Z) multiplet of type IIB superstrings". In: *Physics Letters* B360 (1995), pp. 13–18. DOI: 10.1016/0370-2693(95)01138-G. arXiv: hep-th/9508143 [hep-th].

[323] S. Seehars et al. "Information Gains from Cosmic Microwave Background Experiments". In: *Physical Review* D90.2 (2014), p. 023533. DOI: 10.1103/PhysRevD.90.023533. arXiv: 1402.3593 [astro-ph.CO].

[324] U. Seljak. "Measuring polarization in cosmic microwave background". In: *The Astrophysical Journal* 482 (1997), p. 6. DOI: 10.1086/304123. arXiv: astro-ph/9608131 [astro-ph].

[325] U. Seljak, U-L. Pen, and N. Turok. "Polarization of the microwave background in defect models". In: *Physical Review Letters* 79 (1997), pp. 1615–1618. DOI: 10.1103/PhysRevLett.79.1615. arXiv: astro-ph/9704231 [astro-ph].

[326] U. Seljak and A. Slosar. "B polarization of cosmic microwave background as a tracer of strings". In: *Physical Review* D74 (2006), p. 063523. DOI: 10.1103/PhysRevD.74.063523. arXiv: astro-ph/0604143 [astro-ph].

[327] U. Seljak and M. Zaldarriaga. "A Line of sight integration approach to cosmic microwave background anisotropies". In: *The Astrophysical Journal* 469 (1996), pp. 437–444. DOI: 10.1086/177793. arXiv: astro-ph/9603033 [astro-ph].

[328] U. Seljak and M. Zaldarriaga. "Signature of gravity waves in polarization of the microwave background". In: *Physical Review Letters* 78 (1997), pp. 2054–2057. DOI: 10.1103/PhysRevLett.78.2054. arXiv: astro-ph/9609169 [astro-ph].

[329] R. Serfozo. *Basics of Applied Stochastic Processes*. Probability and Its Applications. Springer Berlin Heidelberg, 2009.

[330] E. P. S. Shellard. "Cosmic Strings Interactions". In: *Nuclear Physics* B283 (1987), pp. 624–656.

[331] R. K. Sheth and G. Tormen. "Large scale bias and the peak background split". In: *Monthly Notices of the Royal Astronomical Society* 308 (1999), p. 119. DOI: 10.1046/j.1365-8711.1999.02692.x. arXiv: astro-ph/9901122 [astro-ph].

[332] X-D. Shi and G. M. Fuller. "A New dark matter candidate: Nonthermal sterile neutrinos". In: *Physical Review Letters* 82 (1999), pp. 2832–2835. DOI: 10.1103/PhysRevLett.82.2832. arXiv: astro-ph/9810076 [astro-ph].

[333] B. Shlaer, A. Vilenkin, and A. Loeb. "Early structure formation from cosmic string loops". In: *Journal of Cosmology and Astroparticle Physics* 1205 (2012), p. 026. DOI: 10.1088/1475-7516/2012/05/026. arXiv: 1202.1346 [astro-ph.CO].

[334] J. Silk. "Cosmic Black-Body Radiation and Galaxy Formation". In: *The Astrophysical Journal* 151 (Feb. 1968), p. 459. DOI: 10.1086/149449.

[335] J. Silk. "Fluctuations in the Primordial Fireball". In: *Nature* 215 (Sept. 1967), pp. 1155–1156. DOI: 10.1038/2151155a0.

[336] T. H. R. Skyrme. "A Non-Linear Field Theory". In: *Proceedings of the Royal Society of London* A260.1300 (1961), pp. 127–138. ISSN: 0080-4630. DOI: 10.1098/rspa.1961.0018.

[337] R. E. Smith et al. "Stable clustering, the halo model and nonlinear cosmological power spectra". In: *Monthly Notices of the Royal Astronomical Society* 341 (2003), p. 1311. DOI: 10.1046/j.1365-8711.2003.06503.x. arXiv: astro-ph/0207664 [astro-ph].

[338] T. L. Smith, E. Pierpaoli, and M. Kamionkowski. "A new cosmic microwave background constraint to primordial gravitational waves". In: *Physical Review Letters* 97 (2006), p. 021301. DOI: 10.1103/PhysRevLett.97.021301. arXiv: astro-ph/0603144 [astro-ph].

[339] G. F. Smoot et al. "Structure in the COBE differential microwave radiometer first-year maps". In: *The Astrophysical Journal Letters* 396 (Sept. 1992), pp. L1–L5. DOI: 10.1086/186504.

[340] J. Sola and H. Stefancic. "Effective equation of state for dark energy: Mimicking quintessence and phantom energy through a variable lambda". In: *Physics Letters* B624 (2005), pp. 147–157. DOI: 10.1016/j.physletb.2005.08.051. arXiv: astro-ph/0505133 [astro-ph].

[341] T. Solorio et al. "An Active instance-based machine learning method for stellar population studies". In: *Monthly Notices of the Royal Astronomical Society* 363.2 (2005), pp. 543–554. DOI: 10.1111/j.1365-2966.2005.09456.x. arXiv: astro-ph/0507527 [astro-ph].

[342] A. Sornborger et al. "The structure of cosmic string wakes". In: *The Astrophysical Journal* 482.1 PART I (1997), pp. 22–32. ISSN: 0004-637X. DOI: 10.1086/304130.

[343] D. N. Spergel et al. "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters". In: *The Astrophysical Journal Supplement* 148 (2003), pp. 175–194. DOI: 10.1086/377226. arXiv: astro-ph/0302209 [astro-ph].

[344] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *The Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[345] A. A. Starobinsky. "Spectrum of adiabatic perturbations in the universe when there are singularities in the inflation potential". In: *Journal of Experimental and Theoretical Physics Letters* 55 (1992), pp. 489–494.

[346] N. Sugiyama. "Cosmic Background Anisotropies in Cold Dark Matter Cosmology". In: *The Astrophysical Journal* 100 (Oct. 1995), p. 281. DOI: 10.1086/192220. eprint: astro-ph/9412025.

[347] R. A. Sunyaev and Y. B. Zeldovich. "Small-Scale Fluctuations of Relic Radiation". In: *Astrophysics & Space Science* 7 (Apr. 1970), pp. 3–19. DOI: 10.1007/BF00653471.

[348] R. A. Sunyaev and Y. B. Zeldovich. "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies". In: *Comments on Astrophysics and Space Physics* 4 (Nov. 1972), p. 173.

[349] R. S. Sutton and A. G. Barto. *Introduction to Reinforcement Learning*. 1st. Cambridge, MA, USA: MIT Press, 1998.

[350] R. Takahashi et al. "Revising the Halofit Model for the Nonlinear Matter Power Spectrum". In: *The Astrophysical Journal* 761 (2012), p. 152. DOI: 10.1088/0004-637X/761/2/152. arXiv: 1208.2701 [astro-ph.CO].

[351] A. Taruya, T. Nishimichi, and S. Saito. "Baryon Acoustic Oscillations in 2D: Modeling Redshift-space Power Spectrum from Perturbation Theory". In: *Physical Review* D82 (2010), p. 063522. DOI: 10.1103/PhysRevD.82.063522. arXiv: 1006.0699 [astro-ph.CO].

[352] A. Taruya et al. "RegPT: Direct and fast calculation of regularized cosmological power spectrum at two-loop order". In: *Physical Review* D86 (2012), p. 103528. DOI: 10.1103/PhysRevD.86.103528. arXiv: 1208.1191 [astro-ph.CO].

[353] L. J. Tassie. "Cosmic strings, superstrings and the evolution of the Universe". In: *Nature* 323.6083 (1986), pp. 40–42. DOI: 10.1038/323040a0.

[354] J. Tauber et al. "The Scientific programme of Planck". In: (2006). arXiv: astro-ph/0604069 [astro-ph].

[355] I. V. Tetko, D. J. Livingstone, and A. I. Luik. "Neural network studies. 1. Comparison of overfitting and overtraining". In: *Journal of Chemical Information and Computer Sciences* 35.5 (1995), pp. 826–833. DOI: 10.1021/ci00027a006.

[356] The Planck Collaboration. *Planck 2015 TT power spectrum*. Plot. 2015. URL: https://wiki.cosmos.esa.int/planckpla2015/index.php/CMB_spectrum_\%26_Likelihood_Code.

[357] M. A. Troxel and M. Ishak. "The Intrinsic Alignment of Galaxies and its Impact on Weak Gravitational Lensing in an Era of Precision Cosmology". In: *Physics Reports* 558 (2014), pp. 1–59. DOI: 10.1016/j.physrep.2014.11.001. arXiv: 1407.6990 [astro-ph.CO].

[358] R. J. Trumpler and H. F. Weaver. *Statistical Astronomy*. Dover, 1962.

[359] N. Turok. "Causality and the Doppler peaks". In: *Physical Review* D54 (1996), pp. 3686–3689. DOI: 10.1103/PhysRevD.54.R3686. arXiv: astro-ph/9604172 [astro-ph].

[360] S. H. H. Tye, I. Wasserman, and M. Wyman. "Scaling of multi-tension cosmic superstring networks". In: *Physical Review* D71 (2005), p. 103508. DOI: 10.1103/PhysRevD.71.10350810.1103/PhysRevD.71.129906,10.1103/PhysRevD.71.103508,10.1103/PhysRevD.71.129906. arXiv: astro-ph/0503506 [astro-ph].

[361] J. Urrestilla and A. Vilenkin. "Evolution of cosmic superstring networks: A Numerical simulation". In: *Journal of High Energy Physics* 02 (2008), p. 037. DOI: 10.1088/1126-6708/2008/02/037. arXiv: 0712.1146 [hep-th].

[362] J. Urrestilla et al. "Degeneracy between primordial tensor modes and cosmic strings in future CMB data from the Planck satellite". In: *Physical Review* D77 (2008), p. 123005. DOI: 10.1103/PhysRevD.77.123005. arXiv: 0803.2059 [astro-ph].

[363] T. Vachaspati and A. Vilenkin. "Formation and evolution of cosmic strings". In: *Physical Review* D30 (10 1984), pp. 2036–2045. DOI: 10.1103/PhysRevD.30.2036.

[364] T. Vachaspati and A. Vilenkin. "Gravitational radiation from cosmic strings". In: *Physical Review* D31 (12 1985), pp. 3052–3058. DOI: 10.1103/PhysRevD.31.3052.

[365] G. de Vaucouleurs. "Recherches sur les Nebuleuses Extragalactiques". In: *Annales d'Astrophysique* 11 (Jan. 1948), p. 247.

[366] L. Verde, P. Protopapas, and R. Jimenez. "Planck and the local Universe: Quantifying the tension". In: *Physics of the Dark Universe* 2 (2013), pp. 166–175. DOI: 10.1016/j.dark.2013.09.002. arXiv: 1306.6766 [astro-ph.CO].

[367] L. Verde et al. "(Lack of) Cosmological evidence for dark radiation after Planck". In: *Journal of Cosmology and Astroparticle Physics* 1309 (2013), p. 013. DOI: 10.1088/1475-7516/2013/09/013. arXiv: 1307.2904 [astro-ph.CO].

[368] A. Vikhlinin et al. "Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints". In: *The Astrophysical Journal* 692 (Feb. 2009), pp. 1060–1074. DOI: 10.1088/0004-637X/692/2/1060. arXiv: 0812.2720.

[369] A. Vilenkin. "Cosmic strings". In: *Physical Review* D24 (8 1981), pp. 2082–2089. DOI: 10.1103/PhysRevD.24.2082.

[370] A. Vilenkin. "Gravitational field of vacuum domain walls and strings". In: *Physical Review* D23 (4 1981), pp. 852–857. DOI: 10.1103/PhysRevD.23.852.

[371] A. Vilenkin. "Gravitational radiation from cosmic strings". In: *Physics Letters* B107.1 (1981), pp. 47 –50. ISSN: 0370-2693. DOI: 10.1016/0370-2693(81)91144-8.

[372] A. Vilenkin and E. P. S. Shellard. *Cosmic Strings and Other Topological Defects*. Cambridge Monographs on Mathematical Physics. Cambridge: Cambridge University Press, 2000.

[373] X-S. Wei. *"Must Know Tips/Tricks in Deep Neural Networks"*. 2015. URL: "http://lamda.nju.edu.cn/weixs/project/CNNTricks/CNNTricks.html" (visited on 10/19/2015).

[374] S. Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[375] S. Weinberg. "Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions". In: *Reviews of Modern Physics* 52 (1980). [Science210,1212(1980)], pp. 515–523. DOI: 10.1103/RevModPhys.52.515.

[376] S. Weinberg. *Cosmology*. Cosmology. OUP Oxford, 2008.

[377] S. Weinberg. "Gauge and global symmetries at high temperature". In: *Physical Review* D9 (12 1974), pp. 3357–3378. DOI: 10.1103/PhysRevD.9.3357.

[378] M. L. Wilson and J. Silk. "On the anisotropy of the cosmological background matter and radiation distribution. I - The radiation anisotropy in a spatially flat universe". In: *The Astrophysical Journal* 243 (Jan. 1981), pp. 14–25. DOI: 10.1086/158561.

[379] E. Witten. "Bound states of strings and p-branes". In: *Nuclear Physics* B460 (1996), pp. 335–350. DOI: 10.1016/0550-3213(95)00610-9. arXiv: hep-th/9510135 [hep-th].

[380] I. H. Witten et al. *Data Mining: Practical Machine Learning Tools and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2016.

[381] M. Wyman, L. Pogosian, and I. Wasserman. "Bounds on cosmic strings from WMAP and SDSS". In: *Physical Review* D72 (2005), p. 023513. DOI: 10.1103/PhysRevD.72.023513,10.1103/PhysRevD.73.089905,10.1103/PhysRevD.72.02351310.1103/PhysRevD.73.089905. arXiv: astro-ph/0503364 [astro-ph].

[382] M. Wyman et al. "$\nu\Lambda$CDM: Neutrinos help reconcile Planck with the Local Universe". In: *Physical Review Letters* 112 (2014), p. 051302. DOI: 10.1103/PhysRevLett.112.051302. arXiv: 1307.7715 [astro-ph.CO].

[383] B. Xu et al. "Empirical Evaluation of Rectified Activations in Convolutional Network". In: *ArXiv e-prints* (May 2015). arXiv: 1505.00853 [cs.LG].

[384] M. Zaldarriaga and U. Seljak. "An all sky analysis of polarization in the microwave background". In: *Physical Review* D55 (1997), pp. 1830–1840. DOI: 10.1103/PhysRevD.55.1830. arXiv: astro-ph/9609170 [astro-ph].

[385] Y. B. Zeldovich. "A hypothesis, unifying the structure and the entropy of the Universe". In: *Monthly Notices of the Royal Astronomical Society* 160 (1972), 1P. DOI: 10.1093/mnras/160.1.1P.

[386] Y. B. Zeldovich, V. G. Kurt, and R. A. Syunyaev. "Recombination of Hydrogen in the Hot Model of the Universe". In: *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki* 55 (July 1968), pp. 278–286.

[387] W. H. Zurek. "Cosmological experiments in superfluid helium?" In: *Nature* 317.6037 (1985), pp. 505–508. DOI: 10.1038/317505a0.